Lecture Notes in Electrical Engineering 515

Qilian Liang
Xin Liu
Zhenyu Na
Wei Wang
Jiasong Mu
Baoju Zhang *Editors*

Communications, Signal Processing, and Systems

Proceedings of the 2018 CSPS Volume I: Communications

Springer

Lecture Notes in Electrical Engineering

Volume 515

Series Editors

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering - quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

- Communication Engineering, Information Theory and Networks
- Electronics Engineering and Microelectronics
- Signal, Image and Speech Processing
- Wireless and Mobile Communication
- Circuits and Systems
- Energy Systems, Power Electronics and Electrical Machines
- Electro-optical Engineering
- Instrumentation Engineering
- Avionics Engineering
- Control Systems
- Internet-of-Things and Cybersecurity
- Biomedical Devices, MEMS and NEMS

For general information about this book series, comments or suggestions, please contact leontina.dicecco@springer.com.

To submit a proposal or request further information, please contact the Publishing Editor in your country:

China

Jasmine Dou, Associate Editor (jasmine.dou@springer.com)

India

Swati Meherishi, Executive Editor (swati.meherishi@springer.com)

Aninda Bose, Senior Editor (aninda.bose@springer.com)

Japan

Takeyuki Yonezawa, Editorial Director (takeyuki.yonezawa@springer.com)

South Korea

Smith (Ahram) Chae, Editor (smith.chae@springer.com)

Southeast Asia

Ramesh Nath Premnath, Editor (ramesh.premnath@springer.com)

USA, Canada:

Michael Luby, Senior Editor (michael.luby@springer.com)

All other Countries:

Leontina Di Cecco, Senior Editor (leontina.dicecco@springer.com)

Christoph Baumann, Executive Editor (christoph.baumann@springer.com)

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, SCOPUS, MetaPress, Web of Science and Springerlink ****

More information about this series at http://www.springer.com/series/7818

Qilian Liang · Xin Liu ·
Zhenyu Na · Wei Wang ·
Jiasong Mu · Baoju Zhang
Editors

Communications, Signal Processing, and Systems

Proceedings of the 2018 CSPS Volume I: Communications

Springer

Editors
Qilian Liang
Department of Electrical Engineering
University of Texas at Arlington
Arlington, TX, USA

Zhenyu Na
School of Information Science and Technology
Dalian Maritime University
Dalian, China

Jiasong Mu
College of Electronic and Communication Engineering
Tianjin Normal University
Tianjin, China

Xin Liu
School of Information and Communication Engineering
Dalian University of Technology
Dalian, Liaoning, China

Wei Wang
College of Electronic and Communication Engineering
Tianjin Normal University
Tianjin, China

Baoju Zhang
College of Physical and Electronic Information
Tianjin Normal University
Tianjin, China

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-13-6263-7 ISBN 978-981-13-6264-4 (eBook)
https://doi.org/10.1007/978-981-13-6264-4

Library of Congress Control Number: 2019930976

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Contents

Mobile Communications

Ad Hoc Networks

Spectrum Optimization

Wireless Communications

The Base Station Cooperation Scheme Based on Dynamic Transmit Antenna for Massive MIMO Systems

Jing Yang(✉) and Chunhua Zhu

College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China
{yangjing, zhuchunhua}@haut.edu.cn

Abstract. To reduce the hardware complexity mainly caused by radio frequency (RF) chains, transmit antenna selection has been a hot topic in the downlink of massive multiple-input multiple-output (MIMO) systems. This paper will discuss the base station (BS) cooperation scheme based on dynamic transmit antenna. Compared with the fixed-rate coded cooperation scheme, the rate-less codes aided cooperation is adopted at the BSs to reduce the complexity. Moreover, the error performances of the proposed BS cooperation scheme based on dynamic transmit antenna under different modulations are analyzed. And the effectiveness of the proposed scheme is also verified through simulation results in this paper.

Keywords: Massive MIMO · BS cooperation · Dynamic transmit antenna · Coded cooperation

1 Introduction

Large-scale multiple-input multiple-output (MIMO), generally referred to as massive MIMO, is an emerging technology that uses hundreds of transmit antennas serving a limited number of destinations in the same time–frequency resource [1]. Therefore, massive MIMO systems have attracted significant attention in the future 5G mobile communication systems. However, the use of large number of antennas at the base station (BS) leads to an increased computational burden and hardware complexity. In fact, radio frequency (RF) chains are accountable for 50–80% of the transmit/receive power consumptions [2]. As a consequence, antenna selection can be seen as an attractive approach to tackle the inherent hardware complexity of massive MIMO systems [3].

Our focus is on the downlink transmissions (i.e., from BSs to user equipment) which is a primary usage mode for the mobile communication. The broadcast nature of wireless transmission enables surrounding BSs to overhear the transmissions. As a result, each BS can help other adjacent BSs in forwarding their messages, named as BS cooperation. Coded cooperation is one of the most important cooperation strategies, which can achieve both spatial diversity and cooperative coding gains [4]. Various coded cooperation schemes based on fixed-rate channel codes, such as Turbo,

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 3–9, 2019.
https://doi.org/10.1007/978-981-13-6264-4_1

low-density parity check (LDPC), etc., have been investigated [5]. However, when the fixed-rate coded cooperation is applied at the BSs in the downlink of massive MIMO system, plenty of channel codebooks have to be remembered both at the transmitter and the receiver, due to the dynamic transmit antenna for different terminals to maximize the energy efficiency. Therefore, how to design the BS cooperation scheme based on dynamic transmit antenna in the downlink of massive MIMO systems is highly important.

Recently, rate-less codes aided cooperation schemes have attracted considerable interests [6]. Unlike other fixed-rate coded cooperation, rate-less codes aided cooperation does not need to be designed for a predetermined rate. So, it can adapt their code rate to the dynamic transmit antenna. There are several works in the literature that have used rate-less codes aided cooperation to achieve high energy efficiency. For example, two rate-less codes aided cooperation schemes in cellular networks have been investigated in [7]. It has been shown that the mutual information accumulation-based cooperative multicast scheme outperforms the energy accumulation-based cooperative multicast scheme in the spectrum and energy efficiency. Therefore, the rate-less codes aided cooperation is applied to BS cooperation in this paper.

The rest of this paper is organized as follows: The system model is given in Sect. 2. In Sect. 3, the proposed BS cooperation scheme for the downlink of massive MIMO systems is depicted, and the simulation results are developed in Sect. 4. Finally, concluding remarks are provided in Sect. 5.

2 System Model

In this section, we consider the downlink of massive MIMO systems equipped with N_t transmit antennas for the main BS, N_r receive antennas for the cooperative BS, and one receive antenna for each user. Without loss of generality, $N_r \leq N_t$ in this paper.

During the broadcast phase, after encoded, M-ary quadrature amplitude modulation (M-QAM) is implemented for the source information at the main BS. When the channel experiences flat fading, the received signal at the cooperative BS is given by

$$\mathbf{y_{mc}} = \sqrt{\mathbf{G_{mc}}}\sqrt{P_m}\mathbf{H_{mc}}\mathbf{s_{mc}} + \mathbf{n_{mc}} \tag{1}$$

where $\mathbf{G_{mc}}$ is the main BS to the cooperative BS (m–c) path loss-related power gain, P_m is the transmit power at the main BS, $\mathbf{H_{mc}}$ is the channel matrix and its $N_r \times N_t$ entries are independent identical distribution (i.i.d) complex circular symmetric Gaussian random variables with zero mean and unit variance, $\mathbf{n_{mc}}$ is the additive white Gaussian noise, $\mathbf{s_{mc}}$ is the transmitted signal at the main BS, while $\mathbf{y_{mc}}$ is the received signal at the cooperative BS.

During the cooperative phase, first, the received signal $\mathbf{y_{mc}}$ is demodulated and decoded at the cooperative BS. Then, the decoded signal at the cooperative BS is re-encoded and re-modulated, before forwarding to the user. Similarly, when the cooperative BS to user (c–u) channel is flat fading, the received signal at the user is expressed as

$$\mathbf{y_{cu}} = \sqrt{\mathbf{G_{cu}}}\sqrt{P_c}\mathbf{H_{cu}}\mathbf{s_{cu}} + \mathbf{n_{cu}} \tag{2}$$

where $\mathbf{G_{cu}}$ is the c–u path loss-related power gain, P_c is the transmit power at the cooperative BS, $\mathbf{H_{cu}}$ is the i.i.d complex circular symmetric Gaussian random variables with zero mean and unit variance, $\mathbf{n_{cu}}$ is the additive white Gaussian noise, $\mathbf{s_{cu}}$ is the retransmitted signal at the cooperative BS, while $\mathbf{y_{cu}}$ is the received signal at the user. Finally, the received signal $\mathbf{y_{cu}}$ is demodulated and decoded at the user.

3 Proposed Scheme

The main task of this paper is the BS cooperation scheme based on dynamic transmit antenna for the downlink of massive MIMO systems. When the fixed-rate coded cooperation is applied at the BSs, plenty of channel codebooks have to be stored both at the transmitter and the receiver. To address this problem, rate-less codes aided cooperation scheme is investigated in this paper.

In this section, the proposed BS cooperation scheme for the downlink of massive MIMO systems is depicted. When the number of the total users is I, the block diagram of the ith $(1 \le i \le I)$ terminal is shown in Fig. 1.

As shown in Fig. 1(a), the source information is $\mathbf{x} = [x_1, x_2, \cdots, x_K]$. After the transmission scheduler and the rate-less codes aided cooperation, the encoded signal is $\mathbf{y_m} = [y_1, y_2, \cdots, y_{N_m}]$ at the main BS. After the M-QAM, the signal $\mathbf{s_{mc}}$ is transmitted to the cooperative BS during the broadcast phase.

It is worth noting that each user has a different channel state information (CSI) in the downlink of massive MIMO systems. To maximize the energy efficiency, the number of optimal transmit antennas is different. So, the cooperative transmit antennas at the BSs are dynamic for each user. Therefore, assume that Ω is the subset of active transmit antennas to maximize the energy efficiency and its cardinality $|\Omega|$ is $T_m(T_c)$ at the main BS (or at the cooperative BS). When the fixed-rate coded cooperation is applied, plenty of codebooks have to be stored both at the transmitter and the receiver. Fortunately, rate-less codes aided cooperation can adaptively adjust the cooperative rate depending on the dynamic cooperation system. Therefore, the rate-less codes aided cooperation is taken up both at the main BS and the cooperative BS for the downlink of massive MIMO systems in this paper as shown in Fig. 1(a) and (b).

Supposed that the signal $\mathbf{y_{mc}}$ is received at the cooperative BS in the time-delay system. First, the received signal $\mathbf{y_{mc}}$ is demodulated and decoded as shown in Fig. 1 (b). In the proposed BS cooperation scheme, the cooperative phase is started, no matter $\mathbf{s_{cu}}$ is correct or not. The reason is that the correct decoded information may be not obtained at the cooperative BS in real systems to meet the requirement of time delay. Even if the decoded signal $\mathbf{s_{cu}}$ is not absolutely correct, we should retransmit the signal $\mathbf{s_{cu}}$ in some situation. For example, in unequal error protection transmission, part of the correct signals may be significant.

When the retransmit signal $\mathbf{y_{cu}}$ is reached at the ith user, the demodulated and decoded are carried out as shown in Fig. 1(c). In brief, the procedure of the proposed BS cooperation scheme is explained as below.

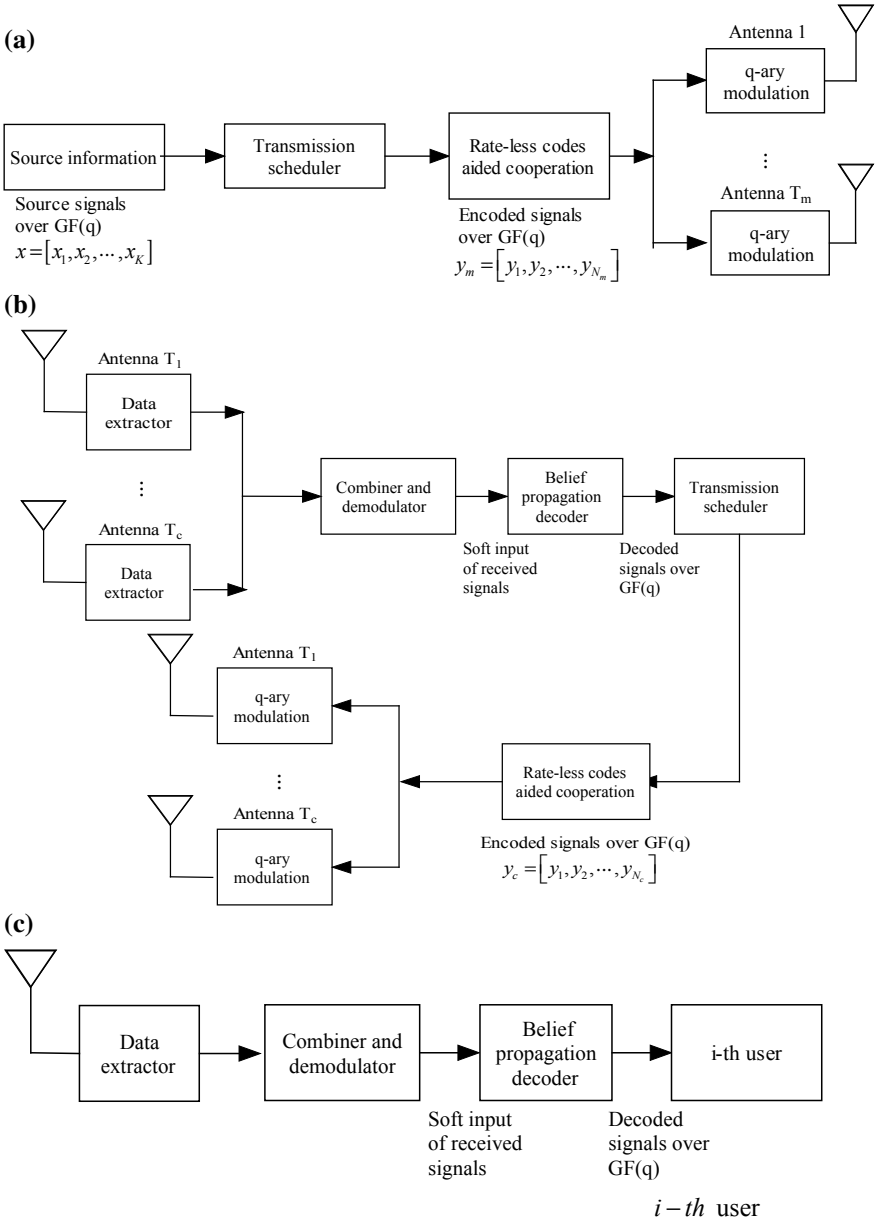


Fig. 1 (a) The block diagram of the proposed scheme at the main BS. (b) The block diagram of the proposed scheme at the cooperative BS. (c) The block diagram of the proposed scheme at the *i*th user.

Step 1: First of all, to maximize the energy efficiency, the number of optimal transmit antennas is supposed to be T_m at the main BS for the ith user. The main task of this paper is the BS cooperation scheme, so the investigation of the concrete antenna selection algorithm is not involved in this paper.

Step 2: In the broadcast phase, the source information is encoded and modulated, before broadcasting to the cooperative BS as shown in Fig. 1(a). Note that the transmitted signal is $\mathbf{s_{mc}}$, the number of the encoded signals is N_m at the main BS in time-delay system, and the number of optimal transmit antennas is T_m.

Step 3: In the cooperative phase, the received signal $\mathbf{y_{mc}}$ is demodulated and decoded at first as shown in Fig. 1(b). No matter the retransmitted signal $\mathbf{s_{cu}}$ is correct or not, re-encoded and re-modulated are adopted in the proposed BS cooperation scheme. Note that the number of the encoded signals is N_c at the cooperative BS, and the number of optimal transmit antennas is T_c.

In real system, different users have different numbers of optimal active transmit antennas to maximize the energy efficiency. Then, the cooperative antenna system is dynamic for each user. As a result, plenty of channel codebooks have to be stored both at the transmitter and the receiver in the fixed-rate coded cooperation scheme. This is unpractical for the portable user in massive MIMO system. Therefore, the rate-less codes aided cooperation is carried out at the BSs for the dynamic transmit antenna in this paper. And the parameters of N_m and N_c are time-varying.

Step 4: Finally, when the signal $\mathbf{y_{cu}}$ is received at the ith user, the demodulation and decoded are implemented, respectively, as shown in Fig. 1(c).

Anyway, the proposed BS cooperation scheme based on dynamic transmit antenna in this paper can reduce the complexity, compared with the fixed-rate coded cooperation scheme both in the broadcast phase and in the cooperative phase.

4 Simulation Results

In this section, we provide simulation results to demonstrate the error performances of the proposed BS cooperation scheme in the downlink of massive MIMO systems. As shown in Fig. 1, the rate-less codes aided cooperation is implemented over the finite field GF(q) with q-ary modulation in the proposed scheme. Then, the symbol conversion from binary to nonbinary at the transmitter and nonbinary to binary for belief propagation (BP) decoding at the receiver are avoided. Generally, rate-less codes aided cooperation can be constructed both from the generator matrix and the parity-check matrix. The example of the former is fountain codes and that of the latter is extended irregular repeat-accumulate (eIRA) codes. It is shown that the eIRA codes have capacity achieving performance over wireless fading channels with lower encoding and decoding complexity [8], which is utilized in this paper. When the source information $\mathbf{x} = [x_1, x_2, \ldots, x_K]$, the jth encoded signal $y_j (1 \leq j \leq N_m)$ or $(1 \leq j \leq N_c)$ at the BSs can be described as

$$y_j = \begin{cases} x_1, & j = 1 \\ y_{j-1} + x_j, & 1 < j \le K \\ y_{j-1} + c_{j,1}x_{j,1} + c_{j,2}x_{j,2} + \cdots c_{j,d_j}x_{j,d_j} & j > K \end{cases} \quad (3)$$

where $c_{j,1}, c_{j,2}, \cdots, c_{j,d_j}$ is the weight of the selected transmit information, $x_{j,1}, x_{j,2}, \cdots, x_{j,d_j}$ is the selected transmit information for the jth encoded symbol. In the proposed scheme, the cooperative BS will retransmit the information no matter it is decoded successfully or not. Then, in our simulations, the parameters of the rate-less codes aided cooperation at the BSs are $d_j = 3$, $K = 500$, $N_m = N_c = 1000$. For simplification, the parameters $\mathbf{G_{mc}}, P_m, \mathbf{G_{cu}}, P_c$ are equal to 1, the number of the optimal transmit antennas T_m at the main BS is 20, and T_c at the cooperative BS is 4/8/12, respectively. In addition, the main BS and the cooperative BS have the same modulation scheme in simulations.

We would like to show the bit error rate (BER) performance of the proposed BS cooperation scheme based on dynamic transmit antenna for different modulations. For example, when Binary Phase Shift Keying (BPSK) is adopted and the number of the optimal transmit antennas T_c is 4 at the cooperative BS, it abbreviates as BPSK-4 as shown in Fig. 2. It can be seen that better performance gain can be achieved if BPSK is adopted, compared with that of 16-ary QAM (16QAM). For example, when the number of optimal transmit antennas T_c is 12 at the cooperative BS, the proposed scheme using BPSK achieves about 3×10^{-4} at the Eb/N0 of 8 dB, while that of the 16QAM is above 10^{-2}, whereas the spectrum effectiveness is higher in 16QAM, contrast to that of BPSK. Anyway, the results of different modulations will help to guide the design of coded cooperation scheme based on dynamic transmit antenna in the future energy-efficient wireless communication systems.

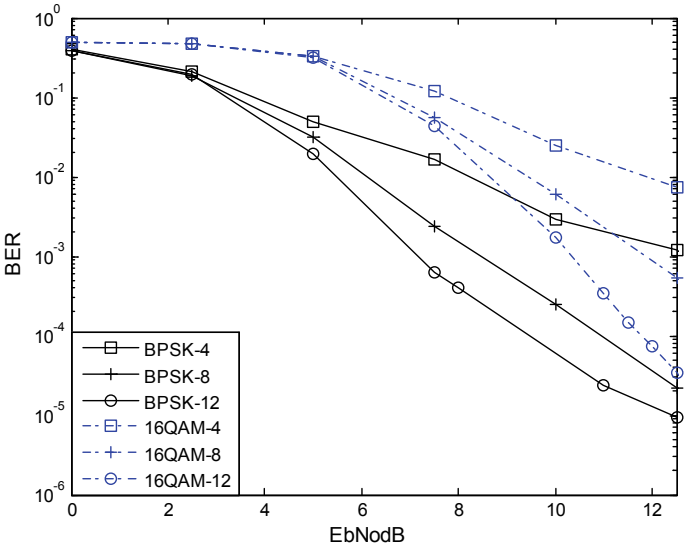


Fig. 2 BER versus Eb/N0 for the different modulations with dynamic transmit antenna.

5 Conclusions

In this paper, we investigate the error performance of the proposed BS cooperation scheme based on dynamic transmit antenna in the downlink of massive MIMO systems where different modulations are taken into account. Compared with the fixed-rate encoded cooperation, the rate-less codes aided cooperation at the BSs is adopted in this paper for the dynamic transmit antenna to reduce the complexity. The proposed BS cooperation scheme gets distinct performance gain for different modulations. For example, when the number of optimal transmit antenna at the cooperative BS is 12, the proposed scheme using BPSK achieves about 3×10^{-4} at the Eb/N0 of 8 dB.

Acknowledgements. This paper was partially supported by the National Natural Science Foundation of China under Grant 61601170, 61741107, and Natural Science Foundation of Henan Provincial Education Department under Grant 17A510001.

References

1. Larsson, E., Edfors, O., Tufvesson, F., Marzetta, T.: Massive MIMO for next generation wireless systems [J]. IEEE Commun. Mag. **52**(2), 186–195 (2014)
2. Amadori, P.V., Masouros, C.: Interference driven antenna selection for massive multi-user MIMO [J]. IEEE Trans. Veh. Technol. **65**(8), 5944–5958 (2016)
3. Jiang, C., Cimini, L.: Antennas selection for energy-efficient mimo transmission [J]. IEEE Wirel. Commun. Lett. **1**(6), 577–580 (2012)
4. Mahyar, S., Li, Y.H., Tian, S., Vucetic, B.: Distributed raptor coding for erasure channels: partially and fully coded cooperation [J]. IEEE Trans. Commun. **61**(9), 3576–3589 (2013)
5. Castura, J., Mao, Y.: Rateless coding for wireless relay channels [C]. In: IEEE International Symposium on Information Theory, pp. 810–814 (2005)
6. Alqahtani, A.H., Sulyman, A.I., Alsanie, A.: Rateless space-time block code for mitigating pilot contamination effects in multi-cell massive MIMO system with lossy links [J]. IET Commun. **10**(16), 2252–2259 (2016)
7. Zhang, G.P., Gu, J., Yang, K., Liu, P.: Performance analysis of two cooperative multicast schemes in cellular networks [J]. Wirel. Pers. Commun. **95**(2), 1317–1331 (2017)
8. Bao, J., Zhan, Y., Yin, L., Lu, J.: Design of efficient joint eIRA-coded MSK modulation systems for space communications [J]. IEEE Trans. Aerosp. Electron. Syst. **48**(2), 1636–1642 (2012)

Practical Implementation of MIMO-FBMC System

Su Hu[1(✉)], Qu Luo[1], Jing Zhang[1], Dan Huang[1], Zilong Liu[2], and Yuan Gao[3]

[1] National Key Laboratory on Communications, University of Electronic Science and Technology of China, 611731 Chengdu, China
husu@uestc.edu.cn

[2] 5G Innovation Centre, Institute of Communication Systems, University of Surrey, GU2 7XH Guildford, UK

[3] Department of Electronic Engineering, Tsinghua University, 100084 Beijing, China

Abstract. FBMC is now a promising technique to achieve high spectral efficiency and strong robustness against dispersive channels. Extensive research has been done on FBMC system but mostly limited to theoretical research. This paper focussed on the prototype design for multiple-input multiple-output filter bank multicarrier communication (MIMO-FBMC) system. A practical testbed design for 4×4 MIMO-FBMC system with two National Instruments PXIe and four universal software-defined radio reconfigurable input/output (USRP-RIO) devices is presented. Details of system design including synchronization, channel estimation, and bit error rate are considered.

Keywords: FBMC · MIMO · USRP · Testbed setup

1 Introduction

Orthogonal frequency-division multiplexing (OFDM) has been researched and developed for wired and wireless communications system for the past two decades. It has gained widespread recognition due to its efficient multicarrier modulation to deal with multipath fading channel, as well as low receiver complexity through simple frequency-domain equalization [1].

However, OFDM also presents some limitations; its robustness against the multipath propagation effect highly relies on the insertion of redundant cyclic prefixes (CP) and therefore is obtained at the price of sacrificed spectral efficiency. On the other hand, an alternative solution based on filter bank theory has recently attracted increased interest, called filter bank multicarrier (FBMC) [2,3]. In FBMC, each subcarrier is modulated with an offset quadrature amplitude modulation (OQAM). This allows FBMC to increase its robustness against carrier frequency offset (CFO) and Doppler spread by designing a pulse shape filter, which is well localized in time and frequency domains. Moreover, FBMC

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 10–18, 2019.
https://doi.org/10.1007/978-981-13-6264-4_2

offers enhanced flexibility, higher spectral efficiency, and better spectral containment in bandwidth-sensitive applications. However, there is always an intrinsic imaginary interference among FBMC subcarriers and symbols as the subcarrier functions are now only orthogonal in the real field. This makes the channel estimation for OFDM/OQAM systems more complicated than that in OFDM systems.

Extensive research has been done on OFDM/OQAM system, such as channel estimation, equalization, and multiple-antenna (MIMO), but mostly limited to theoretical research. Universal software-defined radio (USRP) is commonly used as prototyping tool in industrial or academic areas, and it is easy to set up a prototype to evaluate the system performance [4].

In this paper, two National Instruments (NI) PXIe-1085 devices with four USRP devices are used to evaluate the performance of 4×4 MIMO-FBMC system at 2.5 GHz under realistic channel conditions. The FBMC testbed is designed based on fourth-generation (4G) long-term evolution (LTE) structure. In addition, the complex training sequence decomposition (CTSD)-based channel estimation scheme is applied to the FBMC system. Finally, the bit error ratio (BER) performance of MIMO-FBMC is analyzed in the practical testbed and compared with the theoretical results.

The rest of the paper is organized as follows: Sect. 2 presents the system model of FBMC systems. In Sect. 3, the system setup, testbed hardware, and implementation are discussed in detail. Numerical simulations and experimental results are presented in Sect. 4, followed by the summary of this paper in Sect. 5.

2 FBMC System Model

Considering a FBMC baseband model with M subcarriers, in which the subcarrier spacing is $1/T$ with T being the complex symbol interval. The baseband output signal of OFDM/OQAM is expressed as [2]

$$s(l) = \sum_{n \in \mathrm{Z}} \sum_{m=0}^{M-1} a_{m,n} \underbrace{j^{m+n} e^{j2\pi ml/M} g\left(l - n\frac{M}{2}\right)}_{g_{m,n}(l)}, \tag{1}$$

where $a_{m,n}$ is real-valued OQAM symbol transmitted over the mth subcarrier and at the nth time slot. $g(l)$ is the real-valued prototype filter impulse response with length of $L_g = KM$ and K is the overlapping factor. $g_{m,n}(l)$ is the time–frequency translated version of $g(l)$. The transmultiplexer response of the prototype filter is defined as

$$\varsigma_{m,n}^{p,q} = \sum_{l=-\infty}^{\infty} g_{m,n}(l) g_{p,q}^{*}(l). \tag{2}$$

Noted that all the values of $\varsigma_{m,n}^{p,q}$ are purely imaginary, except at $(m,n) = (p,q)$. This imply that associated subcarrier functions $g_{m,n}(l)$ are only orthogonal in the real field. Even in a given ideal channel without multipath, channel

distortion, and noise, there will be the intrinsic interference in FBMC systems for any $(m,n) \neq (p,q)$.

Commonly, in most of the FBMC literature, the channel is assumed to be frequency flat at each subcarrier and constant over the prototype filter duration. The output of the analysis filter bank (AFB) at the pth subcarrier and qth of OFDM/OQAM symbol can be expressed as

$$y_{m,n} = H_{m,n}a_{m,n} + \underbrace{\sum_{(p,q)\neq(m,n)} H_{p,q}a_{p,q}\zeta_{m,n}^{p,q}}_{I_{m,n}} + \eta_{m,n}, \tag{3}$$

where $H_{m,n}$ is the channel frequency response (CFR) at (m,n)th lattice, and $I_{m,n}$ and $\eta_{m,n}$ denote the associated intrinsic interference and noise components, respectively. Considering an $N_T \times N_R$ MIMO-FBMC system, based on the same assumptions as before, we can write an equation analogous to (3) for ith transmit antenna to the kth receive antenna,

$$y_{m,n}^{k} = \sum_{i=1}^{N_T} \left\{ H_{m,n}^{k,i}a_{m,n}^{i} + \sum_{(m,n)\neq(p,q)} H_{p,q}^{k,i}a_{p,q}^{i}\zeta_{m,n}^{p,q} \right\} + \eta_{m,n}^{k}, \tag{4}$$

where $H_{m,n}^{k,i}$ denotes CFR from the ith transmit antenna to the kth receive antenna, and $\eta_{m,n}^{k}$ denotes the corresponding noise component, respectively.

3 Testbed Setup and Key Methods

3.1 Testbed Hardware

The testbed for a 4×4 MIMO-FBMC system is implemented with two NI PXIe-1085 and four universal software-defined radio reconfigurable input/output (USRP-RIO-2943) devices. NI PXIe-1085 features a high bandwidth and all hybrid backplanes to meet the needs of high-performance test and measurement applications. Moreover, it is equipped with a high-performance Intel-i7 processor with 4-GB random access memory.

USRP-RIO-2943 is built on the reconfigurable I/O architecture of NI LabVIEW (Laboratory virtual instrument engineering workbench). The USRP-RIO-2943 used in the system is a 2×2 MIMO transceivers, and it can be extended to a higher channel number of MIMO system. This delivers a hardware and software solution for real-time and high-performance prototyping of wireless communication systems.

3.2 Testbed Synchronization

To achieve time synchronization between all the USRP-RIOs, an external clock source is used for MIMO-FBMC testbed setup. The use of a 10 MHz external clock source can avoid losing connection; therefore, all the devices are allowed

transmitting simultaneously once locked to the clock, and the received data signal is guaranteed to be aligned.

Training sequences and preamble are used for system synchronization and channel estimation. Two identical Zadoff–Chu (ZC) sequences are used for time and frequency synchronization. Owing to its constant envelope and zero autocorrelation property, the linearity requirements of the RF amplifier can be reduced a lot and the synchronization performance is promoted as well.

At the receiver side, peak detection technique has been applied to each antenna. The received signal will be correlated with the local synchronization sequence, then searching for the peaks above a certain threshold in the correlation results. The correlation is implemented with inverse fast Fourier transform (IFFT)/fast Fourier transform (FFT) module. FFT-based synchronization method offers lower complexity compare to tradition correlation. Moreover, this method is compatible with MIMO-FBMC system and without need for other modules. The FFT-based synchronization can be expressed as

$$Cor\left(\tau\right) = \mathcal{F}^{-1}\{\mathcal{F}(S\left(\tau + n\right)) \cdot \mathcal{F}(S^{*}_{local}\left(n\right))\}, \tag{5}$$

where τ is the synchronous starting position, $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote FFT, IFFT operate, S represents a received signal truncation, and S^{*}_{local} is the local training sequence. The synchronous starting index will be detected with corresponding synchronization peak value.

The carrier frequency offset (CFO) is estimated by using the same training symbols. Assuming that symbol timing offset is well estimated, a phase difference $\pi\varepsilon$ will be introduced between two adjacent sampling points, with ε being the CFO value.

$$\begin{aligned} \hat{\phi} &= \frac{1}{N/2} \sum_{n=0}^{N/2-1} \left[S^{*}\left(n + N/2\right) \cdot S\left(n\right)\right], \\ \hat{\varepsilon} &= \frac{N}{2\pi\left(N/2\right)} \arg\left\{\hat{\phi}\right\}, \end{aligned} \tag{6}$$

where $\hat{\phi}$ is the estimated phase shift, $\hat{\varepsilon}$ is the estimated frequency offset, and $\arg\{\cdot\}$ is the complex argument operate.

3.3 Channel Estimation for MIMO-FBMC

To improve the spectrum efficiency, the CTSD method proposed in [5] has been adopted for the testbed setup. Two identical ZC sequences are cascaded in the time domain forming data preambles. Because only real-valued symbols are allowed, the CTSD method decomposing each complex-valued symbol into real and imaginary parts and sending them over two FBMC symbols. The CTSD preamble structure is shown in Fig. 1.

Let $\mathbf{A} = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_K\}$ be a ZCZ sequence set, each of length $M/2$, i.e.,

$$\mathbf{a}_i = \left[a_i\left(0\right), a_i\left(1\right), \cdots, a_i\left(n\right), \cdots, a_i\left(L-1\right)\right]^T,\ 1 \le i \le K. \tag{7}$$

The time-domain training sequences $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_K\}$ can be formed by cascading two identical sequences from A, i.e.,

$$\begin{aligned} \mathbf{c}_i &= [c_i(0), c_i(1), \cdots, c_i(M-1)]^T \\ &= [a_i(0), \cdots, a_i(L-1), a_i(0), \cdots, a_i(L-1)]^T. \end{aligned} \tag{8}$$

Applying discrete Fourier transform (DFT) to each sequence $\mathbf{c}_i$, the corresponding training sequences set in the frequency domain can be expressed as

$$C_i(m) = \begin{cases} 2\sum_{n=0}^{L-1} a_i(n) e^{\frac{-j2\pi mn}{L}}, m \; is \; even \\ 0, m \; is \; odd \end{cases}. \tag{9}$$

As shown in Fig. 1, $\mathbf{C}_i$ will be first decomposed into two real-valued symbols (i.e., real and imaginary parts) and then sent over transmit antenna i.

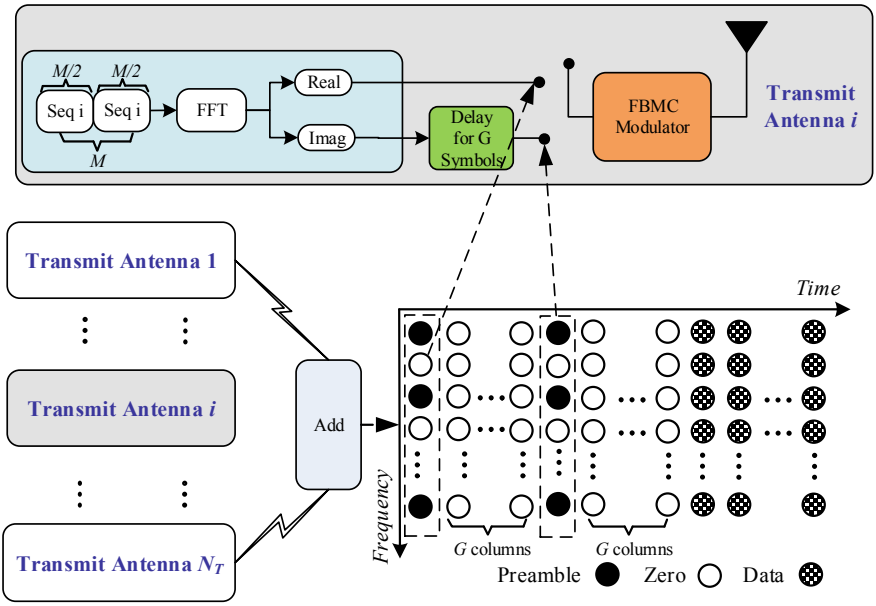


Fig. 1. CTSD preamble structure in MIMO-FBMC system with N_T transmit antennas.

The complex-valued symbols denoted as $\mathbf{y}_k$ are first reconstructed at receiver side, and then applying inverse discrete Fourier transform (IDFT) to $\mathbf{y}_k$, we obtain $\mathbf{r}_k$. By applying linear least square method, the final channel estimator for the kth receive antenna can be expressed as [6]

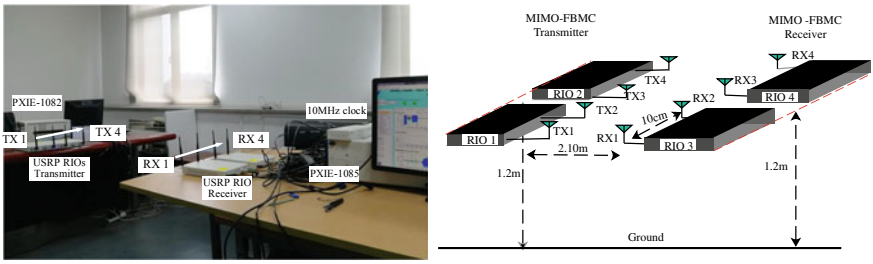
$$\tilde{\mathbf{h}}_k = [\tilde{h}_{1,k}, \tilde{h}_{2,k}, \cdots, \tilde{h}_{N_T,k}]^T = (\mathbf{S}^H\mathbf{S})\mathbf{S}^H\mathbf{r}_k, \tag{10}$$

where $h_{i,k}$ represents the channel impulse response vector from the ith (i = 1, 2, N_T) transmit antenna to the kth(k = 1,2,N_R) receive antenna and $\mathbf{S}_i, i = 1, 2, \cdots, N_T$ is $M \times L_h$ matrices related to $\mathbf{A}$.

4 Experiment Results and Numerical Analysis

4.1 Measurement Camping

The experiment is carried out in indoor laboratory. The experimental setup and its physical layout of 4×4 MIMO-FBMC system are shown in Fig. 2. Four receive antennas are placed about 2.1 m away from the transmit antennas. In addition, the transmit and receive antennas are identical and are placed directly across from each other. All USRP-RIOs are connected to NI PXIe, and the processes of setup synchronization, channel estimation, FBMC/OQAM modulation, demodulation, BER, and SNR calculation are processed in real time.



(a) Experimental setup in laboratory (b) Physical layout of testbed setup

Fig. 2. Experimental setup and Physical layout of 4 × 4 MIMO-FBMC system.

Table 1. System measurement parameters.

Parameters	Values	Parameters	Values
Central frequency	2.5 GHz	System bandwidth	8 MHz
Subcarrier numbers	256	FBMC filter	PHYDYAS, K = 4
Modulation	QPSK/OQAM	Frame length	20 FBMC symbols

The PHYDYAS filter proposed in [7] is used as the prototype filter. The details of testbed parameters are presented in Table 1.

4.2 Experiment Results

The displayed system result of MIMO-FBMC receiver is provided in Fig. 3. The waveform information are located in the middle of Fig. 3, including received

data waveform, synchronization result, and power spectrum density(PSD). As mentioned above, two identical ZC sequences with length of 256 are employed at transmitter side. At the receiver, the local reference sequence is correlated with the received signal and produces corresponding peak value. The synchronization waveform shows that proposed synchronization method works well with good correlation peak by employing IFFT/FFT module.

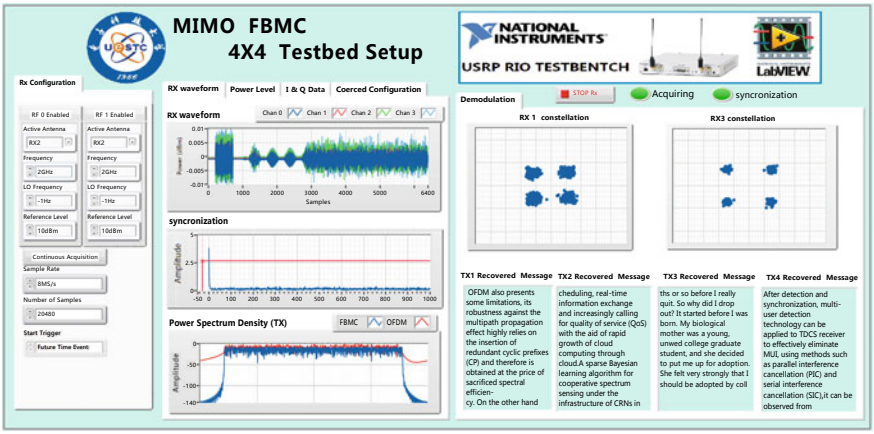


Fig. 3. Displayed results of 4 × 4 MIMO-FBMC system.

The PSD of measured FBMC signal for 256 subcarriers is included in Fig. 3. The PSD of FBMC is compared with OFDM signal, which has the length of cyclic prefix equal to 1/8 FFT length. The figure clearly shows that FBMC signal occupies a lower out-of-band emissions (at least −60 dB) compared to OFDM. FBMC system operates without addition of a guard interval and the out-of-band emissions problem can be well addressed by filtering each subcarrier using a prototype pulse-shaping filter. Moreover, this filter can reduce its side lobes and improve its spectrum efficiency. This advantage provides to the FBMC a good opportunity to substitute the OFDM in next-generation technologies.

The constellations and recovered message at receiver side are shown on the right side of Fig. 3. The received symbols are demodulated online and in real time. The constellation of 10 demodulated FBMC symbols shows that MIMO-FBMC symbols are perfectly reconstructed at the analysis bank by occupying the proposed channel estimate method.

For BER measurements, $10^{\wedge 5}$ symbols were transmitted and calculated. The final BER is averaged among four receive antennas. The averaged BER is taken into compared to the theoretical performance, as shown in Fig. 4. It can be noted that the BER curve of testbed setup experiment result shows slight difference from the MATLAB simulation. This can be attributed to system synchronization problem and the unmeasurable noise from USRP.

For multicarrier transmission scheme, a well-known drawback is that they may suffer from high peak-to-average power ratio (PAPR). The PAPR measurements of OFDM and FBMC signal were performed at receiver side, as shown in Fig. 5. One can see that the FBMC system achieves lower PAPR than OFDM.

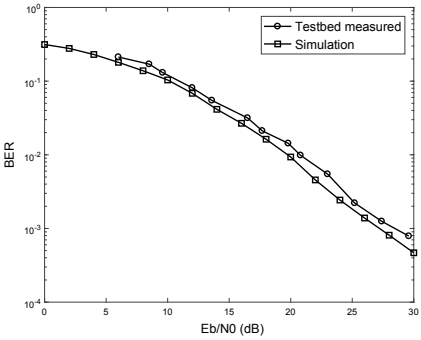


Fig. 4. BER performance of simulation and experiment results.

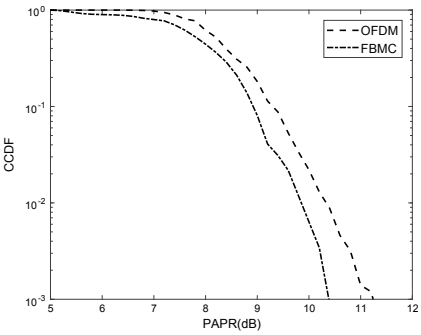


Fig. 5. PAPR comparison between OFDM and FBMC.

5 Conclusion

In this paper, we have presented the prototype design for MIMO-FBMC system by occupying NI-USRP devices. The testbed setup of the transmission link provided a realistic environment, where timing frequency and phase synchronization should be taken into consideration. The practical testbed design has validated that the challenges associated with FBMC, such as MIMO and channel estimation, can be efficiently dealt with. The testbed also gives opportunity to validate more channel estimation methods and complex synchronization algorithms under realistic channel condition.

References

1. Farhang-Boroujeny, B., Kempter, R.: Multicarrier Communication Techniques for Spectrum Sensing and Communication in Cognitive Radios. IEEE Press (2008)
2. Hu, S., Wu, G., Li, T., Xiao, Y., Li, S.: Preamble design with ICI cancellation for channel estimation in OFDM/OQAM system. IEEE Trans. Commun. **93-B**(1), 211–214 (2010)
3. Gao, X., Wang, W., Xia, X., Au, E.K.S., You, X.: Cyclic prefixed OQAM-OFDM and its application to single-carrier FDMA. IEEE Trans. Commun. **59**(5), 1467–1480 (2011)
4. Holland, O., Bogucka, H., Medeisis, A.: 1. The Universal Software Radio Peripheral (USRP) Family of Low-Cost SDRs. Wiley (2015)

5. Hu, S., Liu, Z., Guan, Y.L., Jin, C., Huang, Y., Wu, J.M.: Training sequence design for efficient channel estimation in MIMO-FBMC systems. IEEE Access **PP**(99), 1–1 (2017)
6. Kang, S.W., Chang, K.H.: A novel channel estimation scheme for OFDM/OQAM-IOTA system. ETRI J. **29**(4), 430–436 (2007)
7. PHYDYAS European project. www.ict-phydyas.org

Performance Analysis of Decode-and-Forward Relay in Diffusion Molecular Communication Systems

Jiaxing Wang[1(✉)], Shuo Yuan[1], Wentao Zhou[2], Mahmoud Daneshmand[3], and Mugen Peng[1]

[1] State Key Laboratory of Networking and Switching Technology , Beijing University of Posts and Telecommunications, Beijing, China
jx19882008@163.com, yuanshchn@gmail.com, pmg@bupt.edu.cn
[2] Beijing Bayi School, Beijing, China
tengfei67@sina.com
[3] Stevens Institute of Technology, Hoboken, NJ, USA
mdaneshm@stevens.edu

Abstract. Molecular communication (MC) is a promising paradigm which utilizes molecules to implement communication among nanomachines. Due to short transmit range, decode-and-forward (DF) relay is used to extend the communication distance and achieve reliable remote communications in diffusion-based MC systems, in which the distribution of the number of molecules received at the reception node is modeled as an approximate normal distribution. The impact of DF relay position and the reception radius on the system performance are exploited, and the minimum error detection criterion is used to detect the signal. Our simulation experiments show how the parameters in the model affect the performance of the MC relay system based on diffusion, which can guide us how to improve the performance of the system in the future.

Keywords: Diffusion-based molecular communication · Decode-and-forward (DF) relay · Minimum error probability (MEP) detection

1 Introduction

Molecular communication (MC) is a bio-inspired communication paradigm to enable communication in nano–microscale environment, which has stimulated a great deal of interest in both academic and industrial communities [1]. MC is a cross-research field, which involves a mass of disciplines, such as the biology, computer science, communication engineering, and so on. Potential applications of MC may include health care [2], intelligent drug delivery [3], environment monitoring, military, industry applications, Internet of nanothings (IoNT), etc. Recent interest has concentrated on the design of communication networks among devices that are much smaller than conventional networks.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 19–31, 2019.
https://doi.org/10.1007/978-981-13-6264-4_3

Compared to the traditional electromagnetic wave-based communication, the diffusion-based MC is in nanoscale, and thus has many differences. First, the diffusion-based MC consumes less energy power and lead to higher energy efficiency [4]. Second, the diffusion-based MC employs molecules as the information carrier so it is expected to find a lot of applications in the areas where the operation of conventional electromagnetic-based communication is inefficient or impractical. Finally, the diffusion-based MC has the characteristic of biocompatibility [5], which is suitable to the body area nanonetwork. Molecular communications have received a lot of attention since they are considered an alternative approach to electromagnetic communications due to their unique features.

MC is a novel and different communication paradigm; it considers encoded molecules as information carriers instead of electromagnetic waves as in electromagnetic communications or light waves in optical communications or acoustic waves in acoustic communications. The differences between the general communication and the diffusion-based MC are shown in Fig. 1; the first line represents the traditional communication procedure, and the second line represents the diffusion-based MC procedure; the main distinctions are included in the dotted line frame.

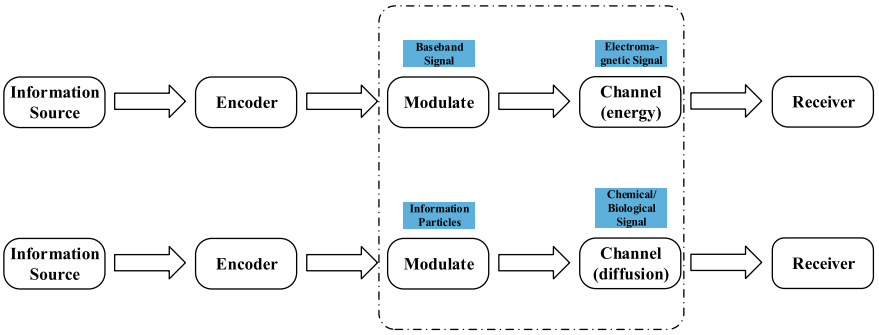


Fig. 1. The general communication and the molecular communication

However, growing the distance will increase the propagation delay and variation of inter-arrival times of molecules, which is a tough challenge for the information retransmission and system synchronization. In addition, due to molecular diffusion and molecular decay, communication range in MC system is strictly limited by the huge channel attenuation. Meanwhile, the stochastic nature of Brownian motion and the unpredictable chemical reaction will produce large channel distortion, which limits the minimum time interval between two messages transmission and creates serious inter-symbol interference (ISI). In order to solve these problems, the relay is used as a bridge in the system. Therefore, some MC relay systems have been proposed in the paper. The sense and forward relaying strategy in a diffusion-based MC is researched in [6], where the molecules are not used as the information carrier, but the bacteria are adopted.

Furthermore, in [7], a DF relaying is researched in a diffusive MC, where a time-dependent MC influenced by both noise and channel memories is modeled, and the impact of emission process and the optimal relay positions in terms of the bit error rate (BER) are evaluated. The paper [8,9] researches the error probability of a two-hop MC for both DF and AF relaying, respectively. In [10], the theoretical system model and the impacts of dielectrophoresis (DEP) relays in diffusive channel in terms of probability of in-sequence delivery of molecules are described. Meanwhile, the information rate of DEP-based channel is analytically obtained for in-sequence. A DF relay-based diffusion MC system including one blood vessel in a human body with positive drift from transmitter to receiver is researched in [11], where the distribution of the number of received molecules with one-dimensional (1-D) environment is normally approximated, and a closed-form expression for the end-to-end BER is derived. The same authors put forward the energy detector applying the diversity combining technique in the cooperative diffusion-based MC network in [12], deriving the closed-form expression for the bit error probability of the proposed detection scheme.

In this paper, a DF relay in diffusion-based MC is researched in a three-dimensional (3-D) manner, and the interference from the previous slot information and channel noise is modeled. The main contributions of this paper are as follows:

- The reception node is modeled as a perfectly absorbing, and the distribution of the number of molecules accumulated at the receiver is obtained by using the approximate normal distribution.
- The closed-form mathematical expression of the error probability performance for the DF relaying in the diffusion-based MC system is derived, which explores the relationship among the position of relay node, the radius of reception, and the channel BER.
- The minimum error probability (MEP) criterion of the standard Bayesian detection framework is applied in the detection part, which can self-adaptive to adjust the detection threshold.

The remainder of this article is organized as follows: The system model for the DF relay in diffusion MC is presented in Sect. 2. The corresponding system performance is analyzed in Sect. 3. The numerical results is presented in Sect. 4. Finally, Sect. 5 gives the conclusions.

2 System Model

In MC, attenuation of the molecular concentration as it travels in the environment via the diffusion process is the major problem. At the steady state, the concentration of molecules is inversely proportional to the distance between the transmitter and the receiver. Therefore, reliable communication to the long distances remains a challenge. Therefore, relay is used to solve this problem and improve the communication range.

The relay in diffusion-based MC system including a transmit node, a relay node, and a receive node is illustrated in Fig. 2. The transmit node is assumed to be a point source. The relay node is located between the transmitter and the receiver; they are in the same line. At the same time, the relay node adopts full-duplex transmission and utilizes DF strategy.

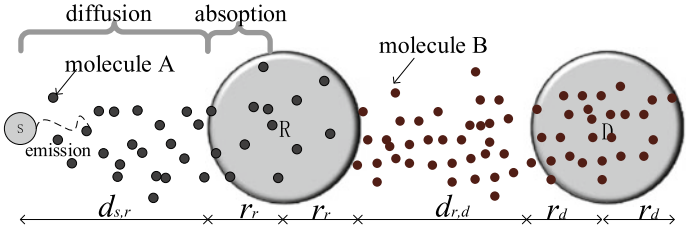


Fig. 2. The diffusion-based MC decode-and-forward relay system.

The symbol duration is defined as the occupied time slots that the transmitter forwards the received information [13], in which the on–off keying (OOK) modulation is used. When the transmitted information is denoted by the bit "1", the transmit node releases Q molecules. When the transmitted information is denoted by the bit "0", the transmit node does not release any molecule. The transmit node uses one certain type of molecules, which can be recognized by the relay node, while the relay node uses another type of molecules, which can be recognized by the receiver. Because two different kinds of molecules are used, the interference can be reduced in this system model. Note that the relay node in MC is different from the traditional relay because the reception and transmission occur simultaneously. Meanwhile, it is assumed that the transmit node, relay node, and the receive node are synchronized [11].

According to [11], the capture time for the concerned 3-D free diffusion in a fluid medium is defined as a released molecule captured by the receiver in the given time t at the first time. The corresponding capture probability density function (PDF) can be expressed as

$$f(t) = \frac{r}{r+d}\frac{1}{\sqrt{4\pi Dt}}\frac{d}{t}e^{-\frac{d^2}{4Dt}}, \tag{1}$$

where d represents the distance from the transmitter to the receiver. r is the coverage radius, and D is the diffusion coefficient. The value of D strictly depends on the molecular type, the environment temperature, the fluid viscosity, and the Stokes' radius of molecules.

The molecules can be captured by the receiver in a given time t, and the cumulative distribution function (CDF) for the capture probability in the absorption time can be expressed as

$$F_{capture}(t) = \frac{r}{d+r}erfc(\frac{d}{\sqrt{4Dt}}), \tag{2}$$

where erfc(.) means complementary error function, and it can be expressed by $erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-u^2} du$. D is the diffusion coefficient of molecules in the medium.

The probability of successful arrival for a molecule transmitted within time t is given by $F(t)$. The transmit node send information bit "1" by emitting a certain number molecules which can be recognized by the relay node at the beginning of the time slot and no molecules to transmit information bit "0". The total number of molecules absorbed by relay node at the n^{th} time slot is denoted by $y_{s,r}[n]$ which consists of three parts. The first part is $C_{Cs,r}[n]$, which means the number of molecules that are sent out and arrived within the current time slot. The second part is $C_{Ps,r}[n]$, which means the number of molecules that are sent at the previous time slot but arrived at the current time slot. Note that these molecules may incur the inter-symbol interference (ISI). The last part is $C_{Ns,r}[n]$, which means the number of molecules that come from the other part of the channel. Hence, the number of molecules received by the relay node can be written as follows [11]:

$$y_{s,r}[n] = C_{Cs,r}[n] + C_{Ps,r}[n] + C_{Ns,r}[n]. \tag{3}$$

As a result of the independent movement of the information molecules with the certain probability of captured by the relay node, $C_{Cs,r}[n]$ is considered as a binominal distribution as [13,14]

$$C_{Cs,r}[n] \sim B(Q_A x[n], F_{r_r,D_A,d_{s,r},t}), \tag{4}$$

where $F_{r_r,D_A,d_{s,r},t} = F_{capture}(r_r, D_A, d_{s,r}, t)$, Q_A represents the transmitter emits the molecular number, and $x[n]$ represents the information bit.

Since the ISI effect after a finite number of previously transmitted symbols becomes negligible, it is assumed that the channel has a finite ISI length. So the $C_{Ps,r}[n]$ obeys

$$C_{Ps,r}[n] \sim B(Q_A x[n-i], F_{i+1s,r} - F_{is,r}), \tag{5}$$

where $x[n-i]$ represents the previous slot information bit, $F_{is,r} = F(r_r, D_A, d_{s,r}, it)$.

It is assumed that $C_{Ns,r}$ obeys the normal distribution as follows:

$$C_{Ns,r}[n] \sim N(\mu_{Ns,r}, \sigma^2_{Ns,r}). \tag{6}$$

Considering the two approximations of the binomial distribution which are Poisson and Gaussian approximations because it is hard to work with the binomial CDF when the consecutively sent symbols, and thus, from (3)–(6), it can derive

$$\begin{aligned} y_{s,r}[n] \sim\ & N(Q_A x[n] F_{1s,r}, Q_A x[n] F_{1s,r}(1 - F_{1s,r})) \\ & + \sum_{i=1}^{I} N(Q_A x[n-i] q^A_{is,r}, Q_A x[n-i] q^A_{is,r}(1 - q^A_{is,r})) \\ & + N(\mu_{Ns,r}, \sigma^2_{Ns,r}), \end{aligned} \tag{7}$$

where $q^A_{is,r} = (F_{i+1s,r} - F_{is,r})$. Hence, $y_{s,r}[n]$ follows the normal distribution as

$$\Pr\left(y^A_{s,r}[n] \,|x^A_s[n] = 0\right) \sim N\left(\mu_{0s,r}, \sigma^2_{0s,r}\right), \tag{8}$$

$$\Pr\left(y^A_{s,r}[n] \,|x^A_s[n] = 1\right) \sim N\left(\mu_{1s,r}, \sigma^2_{1s,r}\right), \tag{9}$$

where the mean and variance are derived as follows:

$$\mu_{0s,r} = p_1 Q_A \sum_{i=1}^{I} q_{is,r} + \mu_{Ns,r}, \tag{10}$$

$$\mu_{1s,r} = p_1 Q_A \sum_{i=1}^{I} q_{is,r} + Q_A F_{1s,r} + \mu_{Ns,r}, \tag{11}$$

$$\sigma^2_{0s,r} = p_1 Q_A \sum_{i=1}^{I} q_{is,r}(1 - q_{is,r}) + p_1 p_0 Q_1^2 \sum_{i=1}^{I} q^2_{is,r} + \sigma^2_{Ns,r}, \tag{12}$$

$$\sigma^2_{1s,r} = Q_A P_{1s,r}(1 - P_{1s,r}) + p_1 Q_A \sum_{i=1}^{I} q_{is,r}(1 - q_{is,r}) + p_1 p_0 Q_A^2 \sum_{i=1}^{I} q^2_{is,r} + \sigma^2_{Ns,r}, \tag{13}$$

where $Pr(x_s[n] = 1) = p_1$ and $Pr(x_s[n] = 0) = p_0$.

The relay node decodes the signal which has received and releases molecules to re-transmit the information. For detection, the relay node motivated to apply the MEP criterion of the standard Bayesian detection framework as [15]

$$y_{s,r} \underset{0}{\overset{1}{\gtrless}} \frac{\frac{\sigma^2_{0s,r} + \sigma^2_{1s,r}}{2}}{\mu_{1s,r} - \mu_{0s,r}} \ln\left(\frac{p_1}{p_0}\right) + \frac{1}{2}(\mu_{0s,r} + \mu_{1s,r}) \equiv \xi. \tag{14}$$

According to the formula (14), the relay node can decode the received information. After decoding, the relay node re-encodes the signal and then forwards it to the receiver; the signal will experience the same attenuation process as the transmitter to the relay node. Similar to the calculation number of A molecules absorbed by the relay node, it can use the same equations to calculate the B molecules absorbed by the receiver at the $(n+1)^{th}$ time slot. The signal can be expressed by $y^B_{r,d}[n+1]$. The distribution of $y^B_{r,d}[n+1]$ is obtained as follows:

$$\Pr\left(y^B_{r,d}[n+1] \,|x^B_s[n+1] = 0\right) \sim N\left(\mu_{0r,d}, \sigma^2_{0r,d}\right), \tag{15}$$

$$\Pr\left(y^B_{r,d}[n+1] \,|x^B_s[n+1] = 1\right) \sim N\left(\mu_{1r,d}, \sigma^2_{1r,d}\right), \tag{16}$$

where the mean value and the variance value are obtained by the formula (10)–(13) with the corresponding parameters Q_2, D_B, $d_{r,d}$, $q_{ir,d}$ instead of the parameters from the transmitter to the relay node Q_1, D_A, $d_{s,r}$, $q_{is,r}$.

3 Performance Analysis

The corresponding bit error probability of the channel from the transmitter to the relay is calculated by

$$P_{e_{s,r}} = p_0 P_{M_{s,r}} + p_1 P_{F_{s,r}}, \tag{17}$$

where $P_{M_{s,r}}$ and $P_{F_{s,r}}$ denote the mis-detection probability and the false probability, respectively. It follows that

$$P_{M_{s,r}} = Q(\frac{\mu_{1s,r} - \xi}{\sqrt{\frac{\sigma^2_{0s,r}+\sigma^2_{1s,r}}{2}}}),\ P_{F_{s,r}} = Q(\frac{\xi - \mu_{0s,r}}{\sqrt{\frac{\sigma^2_{0s,r}+\sigma^2_{1s,r}}{2}}}), \tag{18}$$

where Q(.) denotes the Q-function.

Meanwhile, the MEP criterion of the standard Bayesian detection framework at the receiver can be expressed by

$$y_{r,d} \underset{0}{\overset{1}{\gtrless}} \frac{\frac{\sigma^2_{0r,d}+\sigma^2_{1r,d}}{2}}{\mu_{1r,d} - \mu_{0r,d}} \ln(\frac{p_1}{p_0}) + \frac{1}{2}(\mu_{0r,d} + \mu_{1r,d}) \equiv \xi'. \tag{19}$$

The corresponding bit error probability of the channel from the relay to the receiver is calculated by

$$P_{e_{r,d}} = p_0 P_{M_{r,d}} + p_1 P_{F_{r,d}}, \tag{20}$$

where

$$P_{M_{r,d}} = Q(\frac{\mu_{1r,d} - \xi'}{\sqrt{\frac{\sigma^2_{0r,d}+\sigma^2_{1r,d}}{2}}}),\ P_{F_{r,d}} = Q(\frac{\xi' - \mu_{0r,d}}{\sqrt{\frac{\sigma^2_{0r,d}+\sigma^2_{1r,d}}{2}}}). \tag{21}$$

According to the above description, we can see that the bit error probability of the whole channel can be expressed as

$$P_e = P_{e_{s,r}}(1 - P_{e_{r,d}}) + (1 - P_{e_{s,r}})P_{e_{r,d}}. \tag{22}$$

According to the calculation of the BER, we can get the expression of the corresponding channel maximum transmission rate. The channel transition matrix is

$$\begin{bmatrix} 1 - P_{M_{s,r}} & P_{M_{s,r}} \\ P_{F_{s,r}} & 1 - P_{F_{s,r}} \end{bmatrix} \begin{bmatrix} 1 - P_{M_{r,d}} & P_{M_{r,d}} \\ P_{F_{r,d}} & 1 - P_{F_{r,d}} \end{bmatrix}$$
$$= \begin{bmatrix} (1 - P_{M_{s,r}})(1 - P_{M_{r,d}}) + P_{M_{s,r}} P_{F_{r,d}} & (1 - P_{M_{s,r}})P_{M_{r,d}} + P_{M_{s,r}}(1 - P_{F_{r,d}}) \\ P_{F_{s,r}}(1 - P_{M_{r,d}}) + (1 - P_{F_{s,r}})P_{F_{r,d}} & P_{F_{s,r}} P_{M_{r,d}} + (1 - P_{F_{s,r}})(1 - P_{F_{r,d}}). \end{bmatrix}. \tag{23}$$

Since the transmitted binary information, according to the channel transition probability, the probabilities of receiving 0 and 1 are obtained, respectively, as

$$\begin{aligned} P_y(0) = &\frac{1}{2}[(1 - P_{M_{s,r}})(1 - P_{M_{r,d}}) + P_{M_{s,r}} P_{F_{r,d}}] \\ &+ \frac{1}{2}[P_{F_{s,r}}(1 - P_{M_{r,d}}) + (1 - P_{F_{s,r}})P_{F_{r,d}}], \end{aligned} \tag{24}$$

$$P_y(1) = \frac{1}{2}[(1 - P_{M_{s,r}})P_{M_{r,d}} + P_{M_{s,r}}(1 - P_{F_{r,d}})] + \frac{1}{2}[P_{F_{s,r}}P_{M_{r,d}} + (1 - P_{F_{s,r}})(1 - P_{F_{r,d}})]. \quad (25)$$

From the channel transition matrix, the conditional probabilities can be obtained as

$$P(0|0) = (1 - P_{M_{s,r}})(1 - P_{M_{r,d}}) + P_{M_{s,r}}P_{F_{r,d}}, \quad (26)$$

$$P(1|0) = (1 - P_{M_{s,r}})P_{M_{r,d}} + P_{M_{s,r}}(1 - P_{F_{r,d}}), \quad (27)$$

$$P(0|1) = P_{F_{s,r}}(1 - P_{M_{r,d}}) + (1 - P_{F_{s,r}})P_{F_{r,d}}, \quad (28)$$

$$P(1|1) = P_{F_{s,r}}P_{M_{r,d}} + (1 - P_{F_{s,r}})(1 - P_{F_{r,d}}). \quad (29)$$

From the information theory, the relay channel transmission rate can be calculated as

$$I(X;Y) = H(Y) - H(Y|X), \quad (30)$$

where $H(Y) = -\sum_{i=0}^{1} P_y(i) \log_2 P_y(i)$, $H(Y|X) = -\sum_{j=0}^{1}\sum_{i=0}^{1} P(x = i, y = j) \log_2 P(y = j|x = i)$.

4 Numerical Results

First, the relay can extend the communication range. It is obvious from Fig. 3 that the DF relay can provide a lower bit error probability than the direct channel in the case of the same transmission distance. It can be explained that using the DF strategy in the system can significantly improve the communication performance between the transmit node and the receive node with respect to the direct transmission.

The transmission rate comparing between the direct transmission and the DF transmission is shown in Fig. 4. With the increase of the released molecular number, the transmission rate of the DF relay channel is much higher than the direct link channel. It is proved that the performance of the DF relay channel transmission is better than the direct transmission link in the same transmission distance. This shows that in the diffusion molecular communication system, in order to achieve better transmission performance, it is necessary to add the relay.

According to Fig. 5, we can conclude that the positions of the relay are different, resulting in different channel performances. When the transmitter and the relay release the same number of molecules, the relay node is located at the midpoint between the transmitter and the receiver, and it can obtain the minimum BER. Moreover, the more the number of emission molecules, the lower the BER. When the transmitter releases the number of molecules more than the relay node, the relay node gets closer to the receiver, and it can get the lower BER. On the other hand, if the transmitter releases the number of molecules less than the relay node, the relay node gets closer to the transmitter, and it can get the lower BER.

From Fig. 6, we can know that the results of the channel transmission rate coincide with the BER. When the transmitter and the relay release the same

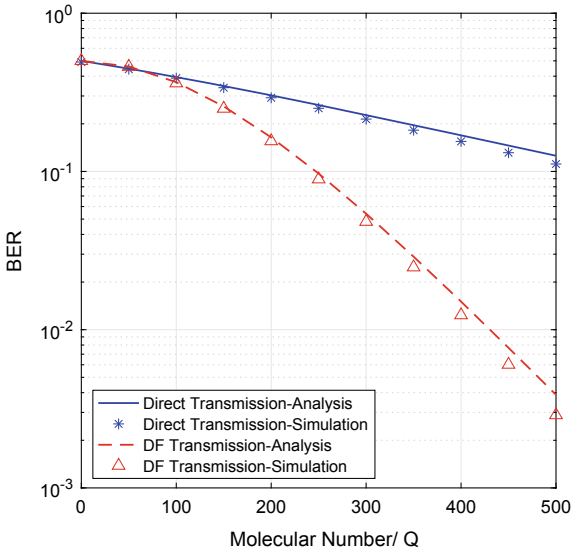


Fig. 3. Bit error probability performance of the diffusion-based MC system with and without the relay node

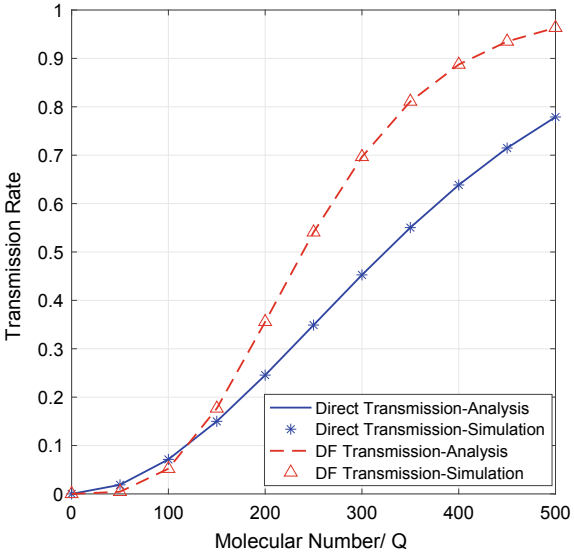


Fig. 4. The transmission rate performance compared between the direct link and the DF relay-based diffusion MC

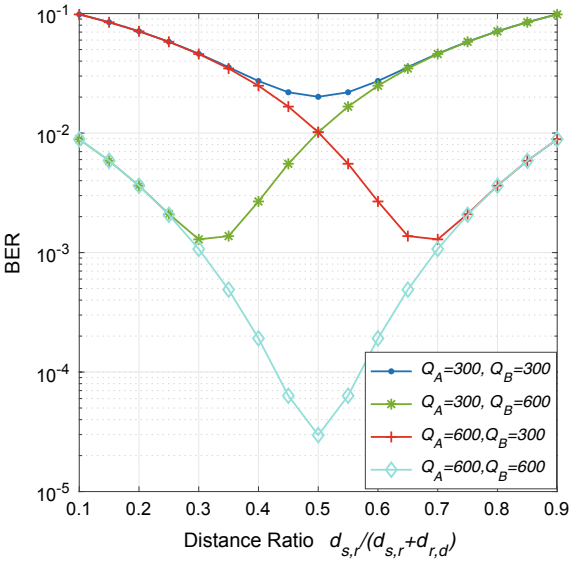


Fig. 5. The bit error rate performance of the DF relay-based diffusion MC with different relay locations

number of molecules, the relay node is located at the midpoint between the transmitter and the receiver, and it can obtain the maximum transmission rate. Moreover, the more the number of emission molecules, the higher the transmission rate. When the transmitter releases the number of molecules more than the relay node, the relay node gets closer to the receiver, and it can get the higher transmission rate. On the contrary, if the transmitter releases the number of molecules less than the relay node, the relay node gets closer to the transmitter, and it can get the higher transmission rate.

The reception radius is also one of the factors that affect the channel performance. From Fig. 7, we can see that when the relay node and the receive node have the same radius, the relay node is located at the midpoint between the transmitter and the receiver, and the BER of channel is minimum. Meanwhile, the larger the reception radius, the smaller the BER. When the reception radius of the relay node is less than the receiver, the relay node should close to the transmitter, which makes the BER of the channel smaller. Otherwise, the relay node should be close to the receiver.

Figure 8 illustrates the position of the relay node from the point of view of the channel transmission rate, which is consistent with the BER. When the reception radius is the same, the relay node is located at the midpoint, and the channel transmission is the maximum.

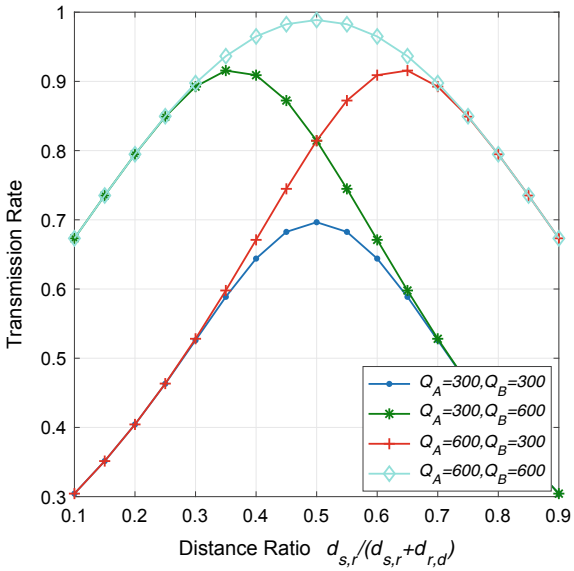


Fig. 6. The transmission rate performance of the DF relay-based diffusion MC with different relay locations

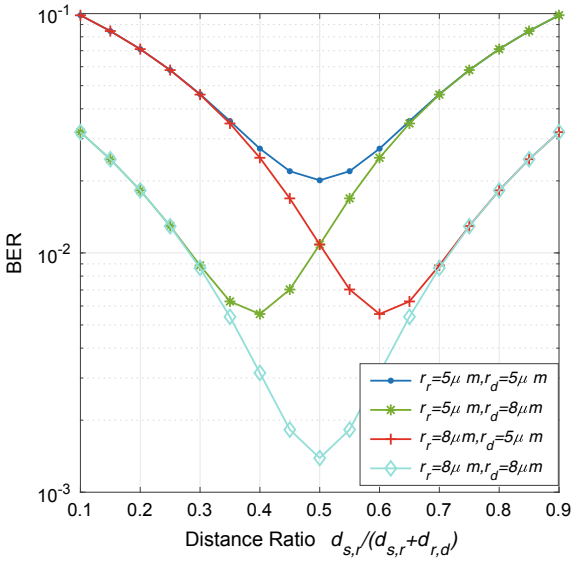


Fig. 7. The bit error rate performance of the DF relay-based diffusion MC with different reception radiuses

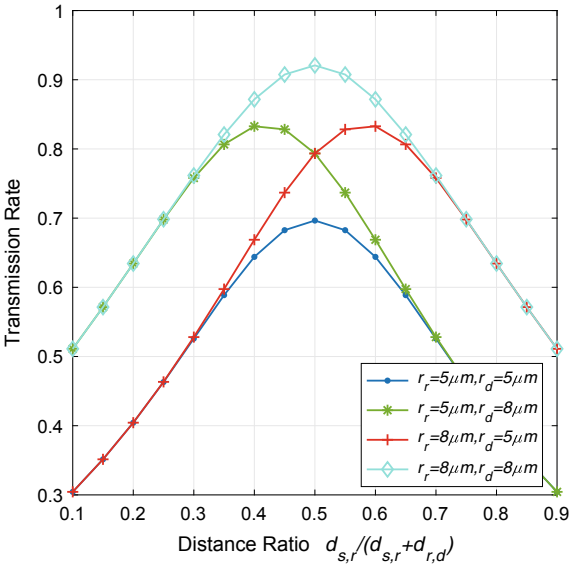


Fig. 8. The transmission rate performance of the DF relay-based diffusion MC with different reception radiuses

5 Conclusion

In this paper, the bit error probability and the transmit rate performances of the decode-and-forward (DF) relay-based diffusion molecular communication (MC) systems have been analyzed and evaluated. A closed-form expression for the bit error probability based on the normal approximation to the distribution of the received signal has been derived. Meanwhile, the minimum error probability (MEP) criterion of the standard Bayesian detection framework is applied in the system, which can self-adaptive to adjust the detection threshold. Simulation results have shown that the performance of the proposed DF relay-based MC is better than the direct transmit within the same transmission distance. In addition, the simulation experiments show how the parameters affect the performance. In the future, we will optimize the system based on the impact of the parameters on the performance of the system.

Acknowledgement. This work was supported in part by the State Major Science and Technology Special Project under Grant 2016ZX03001020-006 and 2017ZX03001025-006, the National Nature Science Foundation of China under No. 61728101, and in part by the National Program for Special support of Eminent Professionals.

References

1. Farsad, N., Yilmaz, H.B., Eckford, A., Chae, C., Guo, W.: A comprehensive survey of recent advancements in molecular communication. IEEE Commun. Surv. Tutor. **18**(3), 1887–1919 (2016)
2. Akyildiz, I.F., Jornet, J.M., Pierobon, M.: Nanonetworks: a new frontier in communications. Commun. ACM **54**(11), 84–89 (2011)
3. Nakano, T., Moore, M.J., Wei, F., Vasilakos, A.V., Shuai, J.: Molecular communication and networking: opportunities and challenges. IEEE Trans. Nanobiosci. **11**(2), 135–148 (2012)
4. Nakano, T., Suda, T., Koujin, T., Haraguchi, T., Hiraoka, Y.: Molecular communication through gap junction channels. Trans. Comput. Syst. Biol. X **54**(10), 81–99 (2011)
5. Nakano, T., Moore, M.: Molecular communication paradigm overview. Next Generation Inf. Technol. **2**(1), 9–16 (2011)
6. Einolghozati, A., Sardari, M., Fekri, F.: Relaying in diffusion-based molecular communication. In: Proceedings of IEEE ISIT 2013, Istanbul, pp. 1844–1848, July 2013
7. Wang, X., Higgins, M.D., Leeson, M.S.: Relay analysis in molecular communications with time-dependent concentration. IEEE Commun. Lett. **19**(11), 1977–1980 (2015)
8. Ahmadzadeh, A., Noel, A., Schober, R.: Analysis and design of two hop diffusion-based molecular communication networks. In: Proceedings of IEEE GLOBECOM, pp. 2820–2825, December 2014
9. Ahmadzadeh, A., Noel, A., Burkovski, A., Schober, R.: Amplify-and-forward relaying in two-hop diffusion-based molecular communication networks. In: Proceedings of IEEE GLOBECOM, pp. 1–7, December 2015
10. Manocha, P., Chandwani, G., Das, S.: Dielectrophoretic relay assisted molecular communication for in-sequence molecule delivery. IEEE Trans. Nanobiosci. **15**(7), 781–791 (2016)
11. Tavakkoli, N., Azmi, P., Mokari, N.: Performance evaluation and optimal detection of relay-assisted diffusion-based molecular communication with drift. IEEE Trans. Nanobiosci. **16**(1), 34–42 (2017)
12. Tavakkoli, N., Azmi, P., Mokari, N.: Optimal positioning of relay node in cooperative molecular communication networks. IEEE Trans. Commun. **16**(1), 34–42 (2017)
13. Singhal, A., Mallik, R.K., Lall, B.: Performance analysis of amplitude modulation schemes for diffusion-based molecular communication. IEEE Trans. Wirele. Commun. **14**(10), 5681–5691 (2015)
14. Kim, N.R., Eckford, A.W., Chae, C.B.: Symbol interval optimization for molecular communication with drift. IEEE Trans. Nanobiosci. **13**(3), 223–229 (2014)
15. Proakis, J.: Digital Communications, 4th edn. McGraw-Hill, New York (2000)

Design and Implementation of a DSSS/UQPSK Demodulation Algorithm Based on FPGA

Yongkui Ma(✉), Shaopeng Zhang(✉), and Jiayan Zhang(✉)

Harbin Institute of Technology, Xidazhijie, 92, Harbin 150001, China
{yk_ma, jyzhang}@hit.edu.cn, spzhang_hit@163.com

Abstract. This paper presents a scheme of UQPSK demodulation algorithm on FPGA which is based on direct sequence spread spectrum (DSSS) communication system. The research is focused on synchronization technology of DSSS/UQPSK signal and design of synchronization loops. The main content of this demodulation algorithm includes carrier synchronization loop and DSSS code synchronization loop. Among them, carrier synchronization loop adopts a second-order feedback closed-loop based on the theory of Costas, and DSSS code synchronization loop uses a method of spread spectrum signal acquisition sliding correlation. In the project that the author participated in, the function of DSSS/UQPSK demodulation algorithm was realized successfully on the FPGA platform.

Keywords: UQPSK · DSSS · Demodulation algorithm · FPGA

1 Introduction

An unbalanced QPSK (UQPSK) is a modulation scheme where the two quadrature components of the carrier can be allocated for unequal power to achieve independent binary data streams of different rates. This modulation scheme is most suitable for transmitting the data streams of two separate sources with varying rates in the same frequency band. In fact, it is the difference in data rates which causes the unequal power ratio when it is desired to have symbol energies and therefore error rates on the two quadrature components to be closer to the same magnitude. This signal design is accomplished in a UAV digital communication link involved in this paper.

The most important advantage of DSSS/UQPSK signal is the accepted transmission security. First, the power spectral density of DSSS is very small due to the extended bandwidth, and the SNR is less than one. It is hard for the blind detection. Second, the two quadrature components of UQPSK use different DSSS codes and powers. It is difficult to detect the two DSSS signals at the same time and calculate their related parameters, respectively [1–3].

Previous results [4, 5] have indicated that when the unbalanced power ratio is large, e.g., approximately 4:1 or greater, a biphase Costas loop is a more efficient demodulator than a fourth-power tracking loop. The Costas loop receiver which tracks on the high data rate signal of the unbalanced QPSK waveform is shown to perform satisfactorily. Approximate error rate computations show that the Costas loop considered performs within a few tenths of a dB of the ideal receiver. In consideration of the

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 32–38, 2019.
https://doi.org/10.1007/978-981-13-6264-4_4

unbalanced power ratio of the designed DSSS/UQPSK signal and the complexity of the system, we choose a biphase Costas loop as the carrier synchronization loop in this paper.

In the following sections, the design and implementation of a DSSS/UQPSK demodulation algorithm based on FPGA are presented. Section 2 describes the DSSS/UQPSK signal model. Section 3 shows the design and implementation of demodulator. Section 4 introduces the results running on FPGA. Finally, Sect. 5 gives the conclusion.

2 DSSS/UQPSK Signal

An unbalanced DSSS/QPSK (UQPSK) signal is defined as [6]

$$s(t) = \sqrt{2P_1}c_1(t)d_1(t)\cos\Phi(t) + \sqrt{2P_2}c_2(t)d_2(t)\sin\Phi(t) \tag{1}$$

where P_1 and P_2 represent the power in the two quadrature components of the carrier, $c_1(t)$ and $c_2(t)$ are two independent DSSS codes, and $d_1(t)$ and $d_2(t)$ are two independent binary data streams, $\Phi(t) = \omega_0 t + \theta(t)$ with ω_0 the radian carrier frequency and $\theta(t) = \theta_0 + \Omega_0 t$ the carrier phase. Input $x(t)$ is immersed in white Gaussian noise

$$x(t) = s(t) + n(t) \tag{2}$$

In (2), $n(t)$ is the additive channel noise which can be expressed in the form of a narrowband process about the actual frequency of the observed data

$$n(t) = \sqrt{2}[N_c(t)\cos\Phi(t) - N_s(t)\sin\Phi(t)] \tag{3}$$

where $N_c(t)$ and $N_s(t)$ are independent baseband noises. Each has single-sided noise spectral density $N_0 \quad W/Hz$, and single-sided bandwidth $B_H < \omega_0/2\pi$.

The input signal $x(t)$ is demodulated by the quadrature reference signals

$$r_c(t) = \sqrt{2}K_1\cos\omega_0 t \quad r_s(t) = \sqrt{2}K_1\sin\omega_0 t \tag{4}$$

Then the baseband phase detector outputs $\varepsilon_c(t)$ and $\varepsilon_s(t)$, neglecting second harmonic terms, are

$$\begin{aligned}
\varepsilon_c(t) &= K_m x(t) r_c(t) \\
&= K_1 K_m\left[\sqrt{P_1}c_1(t)d_1(t) + N_c(t)\right]\cos\theta(t) \\
&\quad + K_1 K_m\left[\sqrt{P_2}c_2(t)d_2(t) - N_s(t)\right]\sin\theta(t) \\
\varepsilon_s(t) &= K_m x(t) r_s(t) \\
&= -K_1 K_m\left[\sqrt{P_1}c_1(t)d_1(t) + N_c(t)\right]\sin\theta(t) \\
&\quad + K_1 K_m\left[\sqrt{P_2}c_2(t)d_2(t) - N_s(t)\right]\cos\theta(t)
\end{aligned} \tag{5}$$

The dynamic error signal $\varepsilon(t)$ is then given by

$$\begin{aligned}\varepsilon(t) =& \frac{1}{2K_1K_m}\varepsilon_c(t)\varepsilon_s(t)\\ =&(\frac{P_1-P_2}{2})sin2\theta(t)\\ &+\sqrt{P_1P_2}c_1(t)c_2(t)d_1(t)d_2(t)\cos 2\theta(t)\\ &+N_c(t)N_s(t)\\ &+\left\{N_c(t)\left[\sqrt{P_1}c_1(t)d_1(t)\sin\theta(t)+\sqrt{P_2}c_2(t)d_2(t)\cos\theta(t)\right]\right.\\ &\left.-N_s(t)\left[\sqrt{P_1}c_1(t)d_1(t)\cos\theta(t)-\sqrt{P_2}c_2(t)d_2(t)\sin\theta(t)\right]\right\}\end{aligned} \tag{6}$$

In (6), from the first term to the last term is, respectively, tracking signal, self-noise, noise, and signal noise. In addition to the desired tracking signal, other terms are considered as equivalent loop noise.

3 The Design and Implementation of Demodulator

The DSSS/UQPSK demodulator architecture is shown in Fig. 1. It should be noted that the carrier frequency of the DSSS/UQPSK signal received from the antenna is GHz magnitude, and the sampling frequency required by the Nyquist theorem is even higher. Considering the existing ADC sampling frequency range and the FPGA processing clock constraints, it is hard to do the digital processing of the input signal directly. The input signal in Fig. 1 is on intermediate frequency after processing.

The DSSS/UQPSK IF signal received is bandpass sampled by the ADC, and the transversion of analog input to full digital processing is completed. Then, the FPGA baseband processing modules are applied to achieve the demodulation algorithm. The two synchronization loops constitute the main part of the demodulator architecture. Among them, the DSSS code synchronization loop contains two parts: the capture loop and the tracking loop. The capture loop achieves the coarse synchronization of the PN code phase, and the tracking loop achieves the fine synchronization on the basis of the capture. The tracking loop is omitted in Fig. 1, and a specific introduction will be made below. The carrier synchronization loop is constructed based on the theory of Costas to achieve the accurate frequency and phase extraction of the carrier.

The carrier synchronization loop and the DSSS code synchronization loop are complementary to each other which enhance the synchronization performance of each other. The DSSS code synchronization loop provides the initial phase of PN code for the carrier synchronization loop, which is used for coherent dispreading. It is obvious that the spread spectrum gain increases the signal-to-noise ratio in the carrier synchronization loop, which is a great help to enhance the performance of carrier synchronization. In the early stage of the system, the local oscillator provides the carrier to the DSSS code synchronization loop with initial frequency offset and phase deviation. After carrier synchronization, the carrier synchronization loop compensates the frequency offset and phase deviation of the local oscillator, so that the DSSS code synchronization loop can extract the data streams correctly.

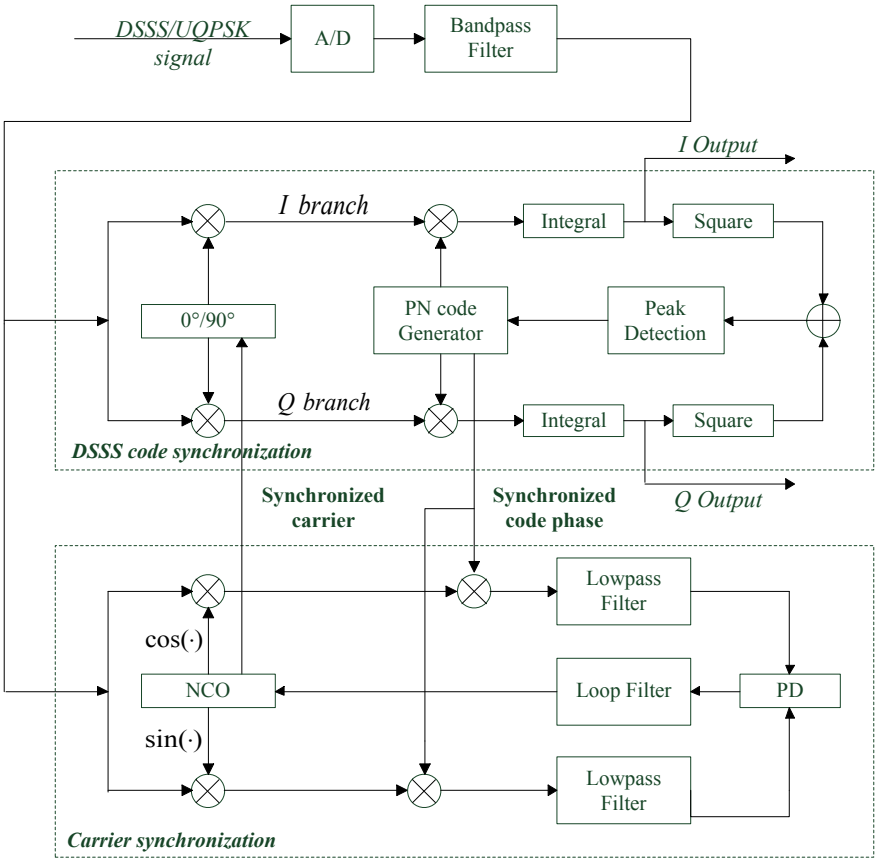


Fig. 1 The DSSS/UQPSK demodulator architecture

3.1 Carrier Synchronization Loop

The DSSS/UQPSK signal is the modulation scheme of suppressing the carrier, and the extraction of its carrier determines the demodulation performance in the system design to a great extent. As introduced in Sect. 1, we choose a biphase Costas loop as the carrier synchronization loop. As introduced in Sect. 2, expression (6) is what we need to process here. Loop filter plays a very important role in the loop, which not only plays the role of low pass filtering but also plays a decisive role in the adjustment of loop parameters. The performance of the ideal two-order phase-locked loop is much better than that of the other loops. In this paper, the ideal two-order loop is used and realized.

The adjustment of the loop parameters is to adjust c_1 and c_2, and the calculation formula of c_1 and c_2 is given directly here [7]

$$\begin{cases} c_1 = \frac{2\xi\omega_n T}{K_d} \\ c_2 = \frac{(\omega_n T)^2}{K_d} \end{cases} \tag{7}$$

where ξ represents loop damping factor, ω_n is loop natural angular frequency, T is frequency word update cycle, and K_d is loop gain. The design of the loop parameters affects capture bandwidth, capture time, and phase jitter. "The capture bandwidth" and "the capture time/phase jitter" are two sets of conflicting parameters. The larger the capture bandwidth is, the greater the phase jitter and capture time is. Therefore, there is a tradeoff and we need to select the suitable parameters for system design.

To reduce the proportion of the logical resource occupancy of FPGA chips and the amount of calculation, decimal shift is used to achieve decimal multiplication. In spite of the problem of effective word length truncation, it can be accepted within the range of satisfying the accuracy of loop error correction.

In addition, combined with the DSSS code synchronization loop, the error lock of carrier can be eliminated. The correlation peak value in DSSS code synchronization loop without Doppler shift is greater than the one with Doppler shift, when the carrier synchronization loop is locked.

3.2 DSSS Code Synchronization Loop

The I branch of the two quadrature components can be allotted with a low data rate and short DSSS code cycle, and therefore high spread spectrum gain is achieved to improve the acquisition sensitivity and meet rapid acquisition required. Besides, the Q branch can be allotted with more power for the data transmission to reduce the BER, when the I branch uses lower power under the premise of measurement accuracy in DSSS code synchronization loop.

Here is a brief introduction to the tracking loop of the DSSS code synchronization loop. The incoherent full-time advance-lag locked loop is adopted as the DSSS code fine synchronization loop. Two independent correlators for correlation operation are used: advanced code correlator and lag code correlator. The architecture of tracking loop is similar to the capture loop, which consists of code correlator, code NCO, loop filter, and code generator. The DSSS code phase demodulation is achieved by code correlator. Then, the input signal is sent to both correlators to achieve correlation operation on advanced local reference code and lag local reference code after multiplied by the local carrier from carrier synchronization loop. Respectively, correlation operation is made on early and late one PN code to obtain the phase of error signal. The phase of error signal running on FPGA using SignalTap is shown in Fig. 2. The phase error signal through the loop filter determines the frequency of code NCO, which makes a dynamic change to control code generator.

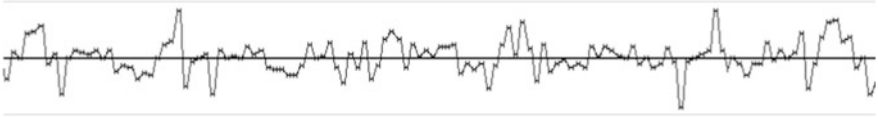


Fig. 2 Phase of error signal running on FPGA

4 Implementation Results

The DSSS/UQPSK demodulation algorithm has been implemented on FPGA, and the results are observed by the SignalTap tool on Stratix II EP2S60F484I4.

The input DSSS/UQPSK signal model is shown in expression (1) in Sect. 2, and the carrier frequency $\omega_0 = 70\,\text{MHz}$. Besides, f_s is the sampling frequency, and $f_s = 99\,\text{MHz}$. Implementation results running on FPGA using SignalTap are shown in Fig. 3.

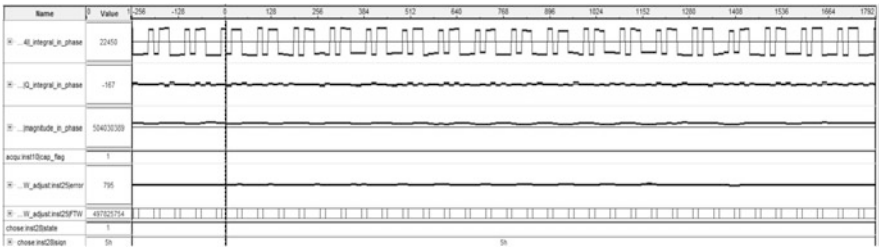


Fig. 3 Implementation results running on FPGA

The demodulated data streams are given here and verified to be in line with expectations. The observed signals of both synchronization loops indicate that the synchronization performance meets expectations.

The logical resource occupancy of FPGA is shown in Table 1.

Table 1 Logical resource occupancy of FPGA

Device	EP2S60F484I4
Timing models	Final
Logic utilization	41%
Combinational ALUTs	4,314/48,352(9%)
Dedicated logic registers	18,974/48,352(39%)
Total registers	18974
Total pins	41/335(12%)
Total virtual pins	0
Total block memory bits	687,488/2,544,192(27%)
DSP block 9-bit elements	6/288(2%)
Total PLLs	2/6(33%)
Total DLLs	0/2(0%)

5 Conclusion

In this paper, we design and implement a DSSS/UQPSK demodulation algorithm on FPGA platform, combining the UQPSK signal model and spread spectrum theory. On the basis of the hardware platform built, the validity and feasibility of this design are verified. Maximum system clock frequency of the demodulation algorithm is 99 MHz, and then the time required to capture the input signal is about an average of 4.2 ms. The implementation results running on FPGA show that the design of the DSSS/UQPSK demodulation algorithm meets the real-time requirements with excellent performance and can be applied to receiver.

This work was supported by the Fundamental Research Funds for the Center Universities (Grant No. HIT.MKSTISP.2016 13).

References

1. Wang, C., Peng, M., Gao, Y.: The study of unbalanced phase shift keying signal modulation recognition algorithm. Signal Inf. Process. **43**(9), 28–32 (2013)
2. Haiyuan, X., Huang, Z., Zhou, Y.: Detection and performance analysis of UQPSK DSSS signals in satellite communication applications. Sig. Process. **24**(3), 361–365 (2008)
3. Shi, S., Zhang, T., Xin, X.: The study of detection of four order cumulants of the UQPSK signal. Sci. Technol. Eng. **14**(20), 257–262 (2014)
4. Weber, C.L.: Candidate receivers for unbalanced QPSK. In: International Telemetering Conference, vol. XII, pp. 455–464 (1976)
5. Levitt, B.K., Lesh, J.R., Springett, J.C.: Shuttle/TDRSS Ku-Band telemetry study. Final Report 900-742, Jet Propulsion Laboratory, Pasadena, California (1976)
6. Braun, W.R., Lindsey, W.C.: Carrier synchronization techniques for unbalanced QPSK signals. In: International Conference on Communications, vol. 1, pp. 1334–1341 (1977)
7. Zhang, A., Du, Y., Han, F.: Design and implementation of costas loop on FPGA platform. In: Electronic Engineer (2006)

Improving Physical Layer Security Using Artificial Noise for Buffer-Aided Relay Networks

Yajun Zhang(✉), Huaming Wang, and Wenli Shi

Army Academy of Artillery and Air Defense (Nanjing Campus), Nanjing, China
yajun2000@foxmail.com, toys123456@sina.com, xiaolizi.shi@163.com

Abstract. It has been verified that cooperative communication technology is an effective way to achieve physical layer security. Recently, people paid much attention to the cooperative networks with buffering relay, which can get a significant performance advantage. In this paper, we investigate the secure problem in decode-and-forward buffer-aided cooperative networks. Based on modified Max-ratio relay selection criteria, artificial noise is transmitted by the source when the selected buffering relay decode and forward message to the destination. The numerical simulation results demonstrate that our proposed Max-ratio with artificial noise (Max-ratio-AN for short) not only outperforms previously reported no-buffer-aided cooperative secure schemes by providing a significant coding gain for small buffer sizes but also ensures a diversity gain enhancement. Comparing to the existing Max-ratio scheme, it also can offer higher secure performance with appropriate artificial noise added.

Keywords: Buffer-aided relay · Physical layer security · Artificial noise · Secrecy outage probability

1 Introduction

Physical layer security, or information-theoretic security, has attracted growing attention recently [1,2]. The basic idea of physical layer security is to exploit physical characteristics of the wireless channel, such as fading or noise, traditionally seen as impediments, to improve the transmission reliability against eavesdropping attacks with relatively low computational overhead. In Wyner's seminal work [3], the maximum nonzero information transmission rate is defined as *secrecy capacity*, at which the transmitter can reliably communicate a secret message to the intended receiver, without the eavesdropper being able to decode it. Since then, many studies have been proposed in the literature from various points [4–7].

Recently, buffer-aided cooperative network has attracted people's attention due to its significant performance advantage compared to the conventional cooperative networks [8]. By introducing data buffers at the relays, it is possible to

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 39–47, 2019.
https://doi.org/10.1007/978-981-13-6264-4_5

relax the constraint that the best source-to-relay and relay-to-destination links for a packet transmission must be determined concurrently. As the pioneering work, [9] first explored two buffering relay models, fixed buffering relay, and dynamic buffering relay, in a three-node relay network, and showed that buffering relay model offers significant performance advantages in terms of capacity over existing relay methods at the expense of increased delay. For multi-buffering-relays cooperative networks, literature [10] proposed Max–Max relay selection (MMRS) scheme. Since the MMRS policy requires a predefined schedule for the relaying transmission (e.g., the second slot is always allocated for relaying), which cause the available diversity degrees not fully exploited, [11,12] examined the more flexible transmission-link selection schemes. Recently, literature [13] proposed *improved Max-weight with link priority* (*imax-weight-priority* for short) scheme, and the outage probability is improved further. You can get more information through literature [14,15].

As the development of buffer-aided relaying technology, using buffering relay to guarantee physical layer security has been proposed in [16–18,20]. In [16,17], the authors consider a half-duplex two-hop relaying network composed of a source (Alice), a destination (Bob), a eavesdropper (Eve), and a decode-and-forward (DF) relay equipped with a buffer. A link selection scheme was proposed through solving two optimization problems, which consider both two-hop transmission efficiency and secrecy constraints. Differing from [16,17], paper [18] investigated multi-relay scenario. The authors propose a novel Max-ratio relay selection scheme for secure transmission in buffer-aided DF relay networks with an eavesdropper, which can intercept signals from both the source and relay nodes. In this paper, we try to use artificial noise (AN) to improve the secure performance of buffer-aided cooperative wireless communication systems. Based on modified Max-ratio relay selection criteria, artificial noise is transmitted by the source when the selected buffering relay decode and forward the message to the destination. To the best of our knowledge, it is the first introducing artificial noise to buffer-aided secure relay system, and the simulation results show the efficacy of proposed Max-ratio-AN scheme.

Notation: Scalar variables are denoted by italic symbols. Vectors and matrices are denoted by lower case and upper case boldface symbols, respectively. $\mathcal{CN}\left(\mu, \sigma^2\right)$ denotes complex Gaussian distribution with mean μ and variance σ^2. The expectation is denoted by $\mathrm{E}\left[\cdot\right]$.

2 System Model and Max-ratio-AN Scheme

The system model of buffer-aided DF relay selection is shown in Fig. 1, where there is one source node (Alice), one destination node (Bob), one eavesdropper (Eve), and K relay nodes ($\mathrm{R}_k, 1 \leq k \leq K$). All nodes operate in the half-duplex mode, which means they do not transmit and receive simultaneously. Each relay is equipped with a data buffer Q_k of finite size L (in the number of data packets). The data packets in the buffer obey the "first-in-first-out" rule.

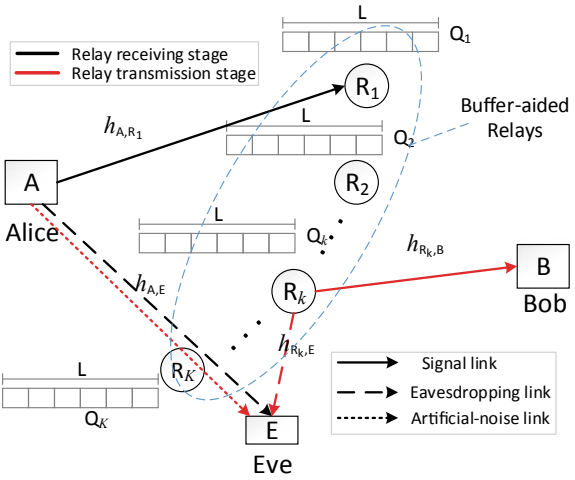


Fig. 1. Buffer-aided relay system model in DF secure transmission, where the eavesdropper intercepts signals from both the source and relay nodes

In our model, we assume no direct link between Alice and Bob due to path loss or shadowing effects. The channel coefficients for A $\rightarrow$ R_k, A $\rightarrow$ E, $\mathrm{R}_k \rightarrow$ B and $\mathrm{R}_k \rightarrow$ E at time t are denoted as $h_{\mathrm{A},\mathrm{R}_k}(t)$, $h_{\mathrm{A,E}}(t)$, $h_{\mathrm{R}_k,\mathrm{B}}(t)$, and $h_{\mathrm{R}_k,\mathrm{E}}(t)$, respectively. We assume that the channels are quasi-static (slow block) Rayleigh fading so that the channel coefficients remain unchanged during one packet duration but independently vary from one packet time to another, i.e., $h_{\mathrm{A},\mathrm{R}_k}(t) \sim \mathcal{CN}(0, \Omega_{\mathrm{A,R}})$, $h_{\mathrm{R}_k,\mathrm{B}}(t) \sim \mathcal{CN}(0, \Omega_{\mathrm{R,B}})$, $h_{\mathrm{A,E}}(t) \sim \mathcal{CN}(0, \Omega_{\mathrm{A,E}})$, and $h_{\mathrm{R}_k,\mathrm{E}}(t) \sim \mathcal{CN}(0, \Omega_{\mathrm{R,E}})$. We assume that all noises are additive white Gaussian noise (AWGN), and without losing generality, the noise variances are all normalized to unity. As the same with some other literature [18], we assume that the exact knowledge of all channels, including the eavesdropping channels, are available.

With the DF applied at the relays, if the relay R^* is the selected buffer-aided relay for the data transmission, the instantaneous secrecy rate for the overall buffer-aided cooperative system is obtained as [18]

$$C_{\mathrm{s}} = \frac{1}{2}\log_2\left(\frac{1+\min\left(\gamma_{\mathrm{A},\mathrm{R}^*}(t), \gamma_{\mathrm{R}^*,\mathrm{B}}(t+\tau)\right)}{1+\gamma_{\mathrm{A,E}}(t)+\gamma_{\mathrm{R}^*,\mathrm{E}}(t+\tau)}\right), \tag{1}$$

where $\gamma_{\mathrm{A},\mathrm{R}^*}(t)$, $\gamma_{\mathrm{A,E}}(t)$, $\gamma_{\mathrm{R}^*,\mathrm{B}}(t+\tau)$ and $\gamma_{\mathrm{R}^*,\mathrm{E}}(t+\tau)$ is the instantaneous SNR of link A $\rightarrow$ R^*, A $\rightarrow$ E, R^* $\rightarrow$ B and R^* $\rightarrow$ E at time t and $(t+\tau)$, respectively.

Here, differing from the existing Max-ratio secure scheme in literature [18], our proposed Max-ratio-AN scheme use a modified Max-ratio relay selection

scheme, which consider the constant parameter of "1+" in the formula (1) as follows:

$$\mathrm{R}^{*}=\arg\max_{\mathrm{R}_k}\left\{\begin{array}{c}\max\limits_{\mathrm{R}_k:\Psi(Q_k)\neq L}\left[\frac{1+\gamma_{\mathrm{A},\mathrm{R}_k}(t)}{1+\gamma_{\mathrm{A},\mathrm{E}}(t)}\right],\\ \max\limits_{\mathrm{R}_k:\Psi(Q_k)\neq 0}\left[\frac{1+\gamma_{\mathrm{R}_k,\mathrm{B}}(t)}{1+\gamma_{\mathrm{R}_k,\mathrm{E}}(t)}\right]\end{array}\right\}, \tag{2}$$

where $\Psi(Q_k)$ gives the number of data packets in buffer Q_k.

Remark 1. Our results in [21] have illustrated the good performance of the proposed Max-ratio relay selection scheme.

Without losing generality, at time slot t, we assume $\mathrm{A}\rightarrow\mathrm{R}^{*}$ is the selected link, and the source Alice transmits data packet $x(t)$ to the relay R^{*} with power P_{A}. The received signal at R^{*} and Eve is given by, respectively,

$$y_{\mathrm{A},\mathrm{R}^{*}}(t)=\sqrt{P_{\mathrm{A}}}h_{\mathrm{A},\mathrm{R}^{*}}(t)\,x(t)+n_{\mathrm{R}^{*}}(t), \tag{3}$$

$$y_{\mathrm{A},\mathrm{E}}(t)=\sqrt{P_{\mathrm{A}}}h_{\mathrm{A},\mathrm{E}}(t)\,x(t)+n_{\mathrm{E}}(t), \tag{4}$$

where $\mathrm{E}\left(|x|^{2}\right)=1$, $n_{\mathrm{R}^{*}}(t)$ and $n_{\mathrm{E}}(t)$ are the noises at the relay R^{*} and eavesdropper E, respectively. Then, $y_{\mathrm{A},\mathrm{R}^{*}}(t)$ is decoded and stored into the buffer $Q_{\mathrm{R}^{*}}$ and waits for its turn to be transmitted.

At the next time slot τ-th, when $y_{\mathrm{A},\mathrm{R}^{*}}(t)$ is decoded and forwarded from R^{*} to the destination node, Alice should transmit artificial noise ω with power P_{J} to confuse eavesdropper Eve, immediately[1]. Then, the received signal at Bob and Eve is given by, respectively,

$$y_{\mathrm{R}^{*},\mathrm{B}}(t+\tau)=\sqrt{P_{\mathrm{R}}}h_{\mathrm{R}^{*},\mathrm{B}}(t+\tau)\,\widehat{x}(t)+n_{\mathrm{B}}(t+\tau), \tag{5}$$

$$\begin{aligned}y_{\mathrm{R}^{*},\mathrm{E}}(t+\tau)=&\sqrt{P_{\mathrm{R}}}h_{\mathrm{R}^{*},\mathrm{E}}(t+\tau)\,\widehat{x}(t)\\&+\sqrt{P_{\mathrm{J}}}h_{\mathrm{A},\mathrm{E}}(t+\tau)\,\omega+n_{\mathrm{E}}(t+\tau)\end{aligned}, \tag{6}$$

where $n_{\mathrm{B}}(t+\tau)$ is the noise for Bob at time $t+\tau$, $\widehat{x}(t)$ is the decode-and-forward data packets of $x(t)$, and its power is P_{R}.

Here, for comparison fairness of different secure schemes, we set the total transmit power constraint to different schemes with the same power budget of P, i.e., $P_{\mathrm{A}}+P_{\mathrm{R}}+P_{\mathrm{J}}=P$. First, we set $P_{\mathrm{J}}=\rho P$, where ρ is power allocation factor. Then, P_{A} and P_{R} can be set as follows:

$$P_{\mathrm{A}}=\frac{(1-\rho)\,P\Omega_{\mathrm{R}^{*},\mathrm{B}}\sigma_{\mathrm{R}^{*}}^{2}}{(\Omega_{\mathrm{R}^{*},\mathrm{B}}\sigma_{\mathrm{R}^{*}}^{2}+\Omega_{\mathrm{A},\mathrm{R}^{*}}\sigma_{\mathrm{B}}^{2})}, \tag{7}$$

$$P_{\mathrm{R}}=\frac{(1-\rho)\,P\Omega_{\mathrm{A},\mathrm{R}^{*}}\sigma_{\mathrm{B}}^{2}}{(\Omega_{\mathrm{R}^{*},\mathrm{B}}\sigma_{\mathrm{R}^{*}}^{2}+\Omega_{\mathrm{A},\mathrm{R}^{*}}\sigma_{\mathrm{B}}^{2})}. \tag{8}$$

[1] Here, perfect synchronization is assumed between Alice and relays. In fact, the synchronization can be easily guaranteed, for example, when Alice has no information data to transmit, she can emit the artificial noise to guarantee the timely jamming confusion for Eve.

Remark 2. It worth noting that the definitions of P_{A} and P_{R} in (7) and (8) are not strictly optimal. Based on the result of literature [22], for DF cooperative scheme, the highest capacity of DF cooperative link can be achieved only if the power allocation between the source and relay can guarantee the same SNRs of relay links. Particularly, P_{A} and P_{R} in (7) and (8) can be derived by the following equation:

$$\begin{cases} \dfrac{P_{\mathrm{A}} \cdot \Omega_{\mathrm{A,R^*}}}{\sigma^2_{\mathrm{R^*}}} = \dfrac{P_{\mathrm{R}} \cdot \Omega_{\mathrm{R^*,B}}}{\sigma^2_{\mathrm{B}}} \\ P_{\mathrm{A}} + P_{\mathrm{R}} = (1-\rho)\, P \end{cases}. \tag{9}$$

Here, we use the average SNRs to calculate P_{A} and P_{R} for decreasing the complexity of realization. It also worth noting that the derivation of P_{A} and P_{R} does not consider the effect of eavesdropper's link, which is an important research direction.

Remark 3. If $\rho = 0$, then $P_{\mathrm{J}} = 0$, which means our proposed Max-ratio-AN scheme can be degraded to Max-ratio scheme in [18]. If $\Omega_{\mathrm{A,R^*}} = \Omega_{\mathrm{R^*,B}}$, then $P_{\mathrm{A}} = P_{\mathrm{R}}$, which means the proposed scheme contains the equal power allocation scenario in [12,18].

3 Secrecy Outage Probability

The secrecy outage probability is defined as the probability that the achievable secrecy rate is less than a target secrecy rate, below which secure transmission is not guaranteed. As such, the secrecy outage probability can be formulated as [1]

$$P_{out}\left(R_{\mathrm{s}}\right) = \Pr\left(C_{\mathrm{s}} < R_{\mathrm{s}}\right) \tag{10}$$

where R_{s} is the target secrecy rate, and C_{s} is expressed in (1).

Based on (3), (4), (5), and (6), the instantaneous SNRs in (1) can be derived by the following formulas:

$$\gamma_{\mathrm{A,R^*}}\left(t\right) = \frac{P_{\mathrm{A}}|h_{\mathrm{A,R^*}}\left(t\right)|^2}{\sigma^2_{\mathrm{R}}}, \qquad \gamma_{\mathrm{A,E}}\left(t\right) = \frac{P_{\mathrm{A}}|h_{\mathrm{A,E}}\left(t\right)|^2}{\sigma^2_{\mathrm{E}}}, \tag{11}$$

$$\gamma_{\mathrm{R^*,B}}\left(t+\tau\right) = \frac{P_{\mathrm{R}}|h_{\mathrm{R^*,B}}\left(t+\tau\right)|^2}{\sigma^2_{\mathrm{B}}}, \tag{12}$$

$$\gamma_{\mathrm{R^*,E}}\left(t+\tau\right) = \frac{P_{\mathrm{R}}|h_{\mathrm{R^*,E}}\left(t+\tau\right)|^2}{P_{\mathrm{J}}|h_{\mathrm{A,E}}\left(t+\tau\right)|^2 + \sigma^2_{\mathrm{E}}}. \tag{13}$$

In Sect. 4, we will investigate the outage performance of the proposed Max-ratio-AN secure scheme by simulation.

4 Simulation Results

In this section, we will simulate the secrecy outage performance of the proposed Max-ratio-AN secure scheme in DF buffer-aided cooperative networks. Though compared to the performance of existing schemes, e.g., T-DFbORS[2], P-DFbORS[3] and Max-ratio, the performance enhancement of our scheme will be examined. Here, with the same assumption of [18], we set $P = 2$.

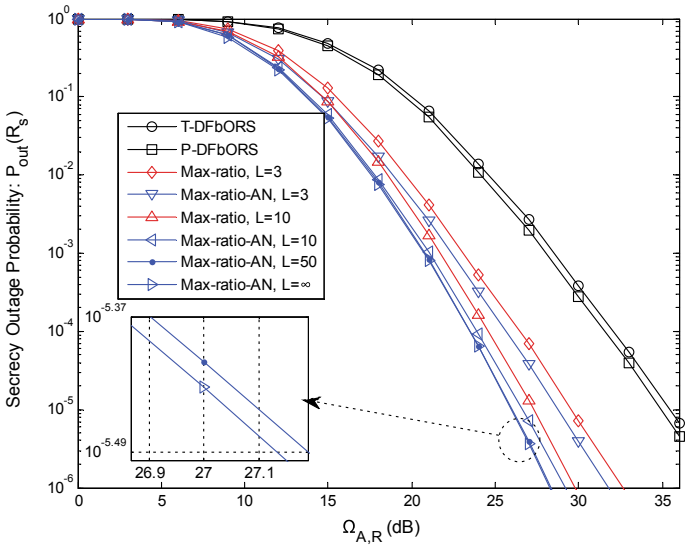


Fig. 2. Secrecy outage probability $P_{out}(R_s)$ v.s. $\Omega_{A,R}$ at $R_s = 1$, $K = 3$, $\rho = 0.3$ for different secure relay selection schemes

Figure 2 compares the secrecy outage performance of the proposed Max-ratio-AN with the exiting T-DFbORS, P-DFbORS, and Max-ratio schemes. In the figure, we set $\Omega_{A,R} = \Omega_{R,B}$, $\Omega_{A,E} = \Omega_{R,E} = 5\,\text{dB}$. We can first see that Max-ratio-AN and Max-ratio have significantly better performance than the other no-buffer-aided two schemes, which shows that the code gain can be obtained through buffering [12]. Moreover, it clearly shown that, for different buffer sizes, the secure performance of our Max-ratio-AN scheme is all better than Max-ratio's, which shows the function of artificial noise. The main reason is that the

[2] T-DFbORS is the abbreviation of the *traditional* DF-based optimal relay selection scheme, it does not consider the wiretap link's channel state information (CSI) and only maximizes the main channel's SNR.

[3] P-DFbORS is the abbreviation of the *proposed* DF-based optimal relay selection scheme [19], it chooses the optimal relay that maximizes the secrecy capacity of DF relaying transmission.

appropriate adding of artificial noise can decrease SNR of the second eavesdropping stage for Eve. Finally, we can observe that, for the proposed Max-ratio-AN scheme, as the buffer size increases, the $P_{out}(R_s)$ would decrease, and slowly approaching the performance bound[4].

Figure 3 shows $P_{out}(R_s)$ versus $\Omega_{A,R}$ for different relay numbers at $R_s = 1$, $L = 3$, and $\rho = 0.3$. In the figure, we also set $\Omega_{A,R} = \Omega_{R,B}$, $\Omega_{A,E} = \Omega_{R,E} = 5\,\text{dB}$. First, we can observe that, for different relay numbers, $P_{out}(R_s)$ of our Max-ratio-AN scheme is all better than Max-ratio's, which also shows the function of artificial noise. Moreover, it is clearly shown that, for the proposed Max-ratio-AN scheme, as the larger of the relays' number, the $P_{out}(R_s)$ would decrease. The main reason is that the larger the L, the bigger the secrecy diversity order[5] can be achieved.

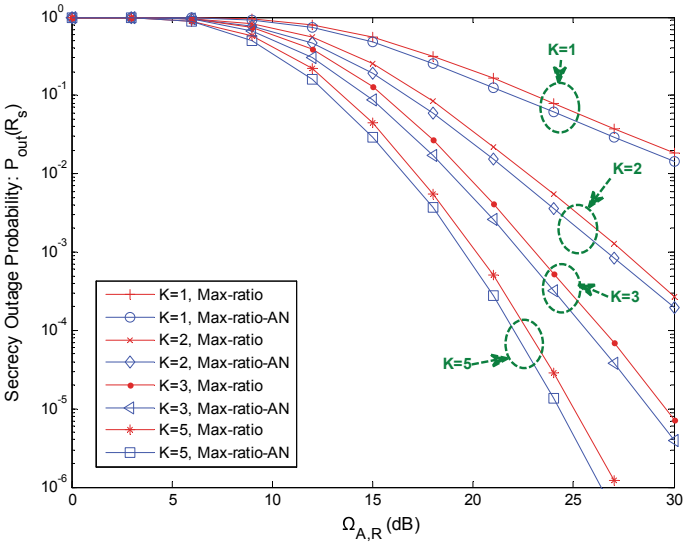


Fig. 3. $P_{out}(R_s)$ v.s. $\Omega_{A,R}$ at $R_s = 1$, $L = 3$, $\rho = 0.3$ for Max-ratio-AN and Max-ratio schemes

5 Conclusion

In this paper, we have proposed Max-ratio-AN scheme for security enhancement in buffer-aided cooperative wireless communication systems. It combined

[4] The theoretic performance bound is an important research work in the future.

[5] Secrecy diversity order determines the slope of the asymptotic outage probability curve [19]. The bigger the value, the outage probability declines more quickly as SNR.

Max-ratio scheme with artificial noise and could offer higher secure performance with appropriate artificial noise added. As the same with Max-ratio scheme, the secrecy outage probability of Max-ratio-AN would be lower when the length of buffer or the number of relay become larger.

Next, analysis for secure performance is an important research work, and the second interesting work is the optimal power allocation between source, relays, and artificial noise.

Acknowledgments. This research was supported by the National Nature Science Foundation of China under grant 61471392 and 41705042.

References

1. Mukherjee, A., Fakoorian, S., Huang, J., Swindlehurst, A.: Principles of physical layer security in multiuser wireless networks: a survey. IEEE Commun. Surv. Tutor. **16**(3), 1550–1573 (2014)
2. Liu, Y., Chen, H.-H., Wang, L.: Physical layer security for next generation wireless networks: theories, technologies, and challenges. IEEE Commun. Surv. Tutor. **19**(1), 347–376 (2017)
3. Wyner, A.D.: The wire-tap channel. Bell Syst. Tech. J. **54**(8), 1355–1387 (1975)
4. Bassily, R., Ekrem, E., He, X., Tekin, E., et al.: Cooperative security at the physical layer-A summary of recent advances. IEEE Signal Process Mag. **30**(5), 16–28 (2013)
5. Hong, Y.W.P., Pang-Chang, L., Kuo, C.C.J.: Enhancing physical-layer secrecy in multiantenna wireless systems: an overview of signal processing approaches. IEEE Signal Process Mag. **30**(5), 29–40 (2013)
6. Harrison, W.K., Almeida, J., Bloch, M.R., McLaughlin, S.W., Barros, J.: Coding for secrecy: an overview of error-control coding techniques for physical-layer security. IEEE Signal Process Mag. **30**(5), 41–50 (2013)
7. Chen, X., Ng, D.W.K., Gerstacker, W.H., Chen, H.H.: A survey on multiple-antenna techniques for physical layer security. IEEE Commun. Surv. Tutor. **19**(2), 1027–1053 (2017)
8. Zlatanov, N., Ikhlef, A., Islam, T., Schober, R.: Buffer-aided cooperative communications: opportunities and challenges. IEEE Commun. Mag. **52**(4), 146–153 (2014)
9. Bing, X., Yijia, F., Thompson, J., Poor, H.V.: Buffering in a three-node relay network. IEEE Trans. Wirel. Commun. **7**(11), 4492–4496 (2008)
10. Ikhlef, A., Michalopoulos, D.S., Schober, R.: Max-max relay selection for relays with buffers. IEEE Trans. Wirel. Commun. **11**(3), 1124–1135 (2012)
11. Lianghui, D., Meixia, T., Fan, Y., Wenjun, Z.: Joint scheduling and relay selection in one- and two-way relay networks with buffering. In: IEEE International Conference on Communications, ICC 2009, Dresden, Germany, pp. 1–5 (2009)
12. Krikidis, I., Charalambous, T., Thompson, J.S.: Buffer-aided relay selection for cooperative diversity systems without delay constraints. IEEE Trans. Wirel. Commun. **11**(5), 1957–1967 (2012)
13. Raza, W., Javaid, N., Nasir, H., Alrajeh, N., Guizani, N.: Buffer-aided relay selection with equal-weight links in cooperative wireless networks. IEEE Commun. Lett. **22**(1), 133–136 (2018)

14. Nomikos, N., Charalambous, T., Krikidis, I., Skoutas, D.N., Vouyioukas, D., Johansson, M., et al.: A survey on buffer-aided relay selection. IEEE Commun. Surv. Tutor. **18**(2), 1073–1097 (2016)
15. Qiao, D.Q., Gursoy, M.C.: Buffer-aided relay systems under delay constraints: potentials and challenges. IEEE Commun. Mag. **55**(9), 168–174 (2017)
16. Jing, H., Swindlehurst, A.L.: Wireless physical layer security enhancement with buffer-aided relaying. In: Asilomar Conference on Signals, Systems and Computers, ACSSC 2013, pp. 1560–1564. CA, USA (2013)
17. Huang, J., Swindlehurst, A.L.: Buffer-aided relaying for two-hop secure communication. IEEE Trans. Wirel. Commun. **14**(1), 152–164 (2015)
18. Gaojie, C., Zhao, T., Yu, G., Zhi, C., Chambers, J.A.: Max-ratio relay selection in secure buffer-aided cooperative wireless networks. IEEE Trans. Inf. Forensics Secur. **9**(4), 719–729 (2014)
19. Yulong, Z., Xianbin, W., Weiming, S.: Optimal relay selection for physical-layer security in cooperative wireless networks. IEEE J. Sel. Areas Commun. **31**(10), 2099–2111 (2013)
20. Liao, X., Wu, Z.Y,Z., Shen, Y., Jiang, X., Inamura, H.: On security-delay trade-off in two-hop wireless networks with buffer-aided relay selection. IEEE Trans. Wirel. Commun. **17**(3), 1893–1903 (2017)
21. Yajun, Z., Tao, L., Aiwei, S.: A New max-ratio relay selection scheme in secure buffer-aided cooperative wireless networks. Electron. Lett. (to be published)
22. Jiangyuan, L., Petropulu, A.P., Weber, S.: On cooperative relaying schemes for wireless physical layer security. IEEE Trans. Signal Process. **59**(10), 4985–4997 (2011)

An Optimized L-Band Single-Tube RF Amplifier

Yuan Gao(✉)

Teacher Education Experimental Teaching Center, Tianjin Normal University, Tianjin 300387, China
284480150@qq.com

Abstract. High linearity power amplifier gain and efficiency is a contradictory issue. It has a high efficiency but the linearity is not ideal. In this dissertation, using the third generation of semiconductor material transistor and studying the structural characteristics of the periphery of each circuit make the efficiency, bandwidth, and linearity get a good result. The main work of this dissertation is to simulate the design of RF power amplifier which is based on GaN HEMT that is produced by Cree Company. First, the dissertation introduced the categories of RF and microwave circuits, the basics of the power amplifier, some necessary parameter, and linearization techniques. Then using HEMT, the single-tube RF high power amplifier is designed which has working bandwidth on L-band. And the output power is more than 50 dBm in the whole bandwidth and the power gain is not less than 14 dB.

Keywords: L-band · RF power amplifier · Wideband

1 Introduction

Radio frequency power amplifier is a very important component in wireless transmitter, which is widely used in wireless remote communication, navigation and positioning, and satellite communication system. Such performances of it such as working bandwidth, output power, added efficiency, and degree of linearity strongly influence the quality of the wireless communication system [1, 2]. Wireless transmitter consists of signal source, the front small signal amplifier, mixer, RF oscillator, RF high power amplifier, and the antenna, as shown in Fig. 1. RF power amplifier is positioned at the final stage of the wireless transmitting system, so the module plays a crucial role; once there is only a slight fluctuation of the index of RF power amplifier, the entire wireless communication system can be affected [3, 4]. Therefore, it is necessary to design one radio frequency power amplifier of a small volume with wider working bandwidth, bigger output power, higher added efficiency, and better linearity.

As higher requirements are put forward regarding wireless communication system, more and more researchers both at home and abroad are making researches and explorations on GaN HEMT devices [5, 6]. China is relatively late in the studies of semiconductor devices of the third generation, but still has made a certain degree of progress.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 48–56, 2019.
https://doi.org/10.1007/978-981-13-6264-4_6

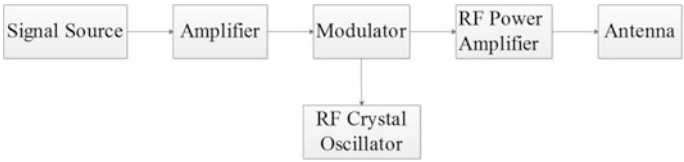


Fig. 1 Block diagram of wireless communication transmitter

In this paper, the high-electron-mobility transistor of Cree Company is used to design a high power amplifier that is based on Range L, and conduct the simulation on the ADS platform [7]. When the RF power amplifier is being designed, DC test is implemented on the power tube first to determine its quiescent operation point; then the biasing circuit and the stable circuit are designed to ensure that power tube stays in absolute stability within the whole frequency range, in case of such serious consequences as auto-oscillation; techniques like source pull and load pull are used to design input and output impedance; finally the combination of microstrip line and capacitor is adopted to design broadband impedance matching network [8].

Other parts of the contents are arranged as follows. Section 1 is the introduction, whose main contents are the research background and the research significance of the subject. Section 2 is about the classification of both RF power amplifier and linearization techniques. Section 3 is about the design framework of single-tube RF high power amplifier that is based on the frequency range of L, and the design of input and output matching network as well as the simulation test on it; Sect. 4 is about the index tests of RF power amplifier, which include parameter test and harmonic test. Section 5 is the summarization of the research design and the planning for the future research direction.

2 RF Power Amplifier

RF power amplifier (PA) is mainly divided into two categories. The traditional power amplifier is under the first category, which is subcategorized into A, AB, B, and C based on the angle of the sine signal during its breakover in transistor [9]. These amplifiers are with high linearity, but its efficiency has to be enhanced. Switching power amplifiers fall into the second category, which is subdivided into D, E, and F during the operation of the transistor [10]. These amplifiers are with high efficiency, but their linearity is relatively low.

2.1 Switching Power Amplifiers

In traditional power amplifiers, the size of the conduction angle is continuously reduced to improve the efficiency of power amplifiers, but there is a limit to the size of the conduction angles. If the conduction angle is too narrow, the output power of the moment is near zero although efficiency can reach 100% theoretically. Class D, Class E, and Class F fall into the category of switching power amplifiers, the sizes of whose conduction angles are all. Its drain efficiency reaches 100% in theory [11].

However, they have to work under the nonlinear state, which leads to unsatisfactory linearity. The structure of dual push–pull output is adopted in Class D power amplifiers, whose long switching time causes the internal loss of transistors under the circumstance of high input signal frequency, so it is not suitable for such high-frequency range as RF [12]. The single-tube switching structure is adopted in Class E and Class F power amplifiers, which avoids disadvantages and brought out the relatively longer conversion time with regard to the signal cycle.

2.2 Linearization Technique

Efficiency and linearity are two conflicting indicators in power amplifier design; linear power amplifier can achieve high linearity with low efficiency, while switching power amplifier can reach high efficiency with unsatisfactory linearity. Therefore, under the premise of high efficiency, it is necessary to adopt the linearization technology to raise the output signal linearity of high power amplifiers. Currently, there are mainly three kinds of linearization techniques, which are power back-off (with the reduction of input power, the working power amplifier stays in linear state, but is kept away from the saturation zone; no extra auxiliary circuits are needed while the efficiency of power amplifiers is weakened [13]; moreover, when IMD3 is below −40 dB, it is hard to improve the linearity of the output signal); feedforward (two ring circuits are used to eliminate the carrier components in output signal and gain the distortion components, and to offset the distortion components of the original output signal with the gained distortion components, so as to achieve linearity enhancement [14]; under the condition of good matching between the amplitude and the phase, the effect of linearity improvement is terrific [15], so this approach requests that the extracted error should be the same as the original signal in amplitude, and the phase difference should be kept at 180° [16]); pre-distortion (one circuit is added in the front end of the power amplifier, which gives the input signal to the corresponding nonlinear inter-modulation components opposite to the nonlinear components produced by the power amplifier devices; thus the nonlinear characteristic of the amplifier is compensated; pre-distortion is subcategorized into digital pre-distortion and simulation pre-distortion, both of whose principles are about insertion of a pre-distorter between the input signal and nonlinear RF power amplifiers, which compensates the amplitude–amplitude distortion (AM–AM) and amplitude–phase distortion (AM–PM) produced by the nonlinear RF power amplifiers, in which the linear relation is established between input and output of the whole pre-distortion system eventually).

3 Design of Single-Tube RF High Power Amplifier of Range L

This section mainly puts forward a design plan of single-tube RF high power amplifier of range L, and the simulation verification is carried out on the plan. The CGH40120F GaN high-electron-mobility transistor (HEMT) of Cree Company is used in the RF power amplifier. One 28 V rail is used for CGH40120F, which offers high efficiency, high gain, and broad frequency band, and suits applications like radio frequency.

RF power amplifier is designed with CGH40120F used on ADS2009 platform, whose parameters are as follows:

1. working bandwidth: 950 MHz–1250 MHz,
2. output power: $\geq$100 W,
3. power gain: $\geq$14 dB,
4. auxiliary efficiency: $\geq$45%, and
5. plainness: $\leq \pm$0.1 dB.

3.1 Idea of RF Power Amplifier

Different from low-frequency circuit, the working frequency range, bandwidth, and other special factors are put into consideration in the design of RF power amplifier, especially the matching between incoming voltage wave and the incoming current wave and active devices, so as to minimize voltage standing wave ratio (VSWR). The first step of RF power amplifier design is to set the static working point of power tube and to build a biasing circuit; then stability analysis is conducted, to gain an absolute stability condition for the amplifier, and input and output impedance of the power tube is measured through the simulation software, in which the matching network between input and output is gained through design, as shown in Fig. 2.

Fig. 2 Steps for design of RF power amplifier

3.2 Design of Biasing Circuit

Before the static working point and the design biasing circuit are determined, the working state of the RF power amplifier needs to be determined first. Comprehensively speaking, when the designed RF power amplifier stays under the working state of Class AB, its efficiency and linearity are relatively better. Because of different static working points, the large signal characteristics of the power tube are totally different from its small signal characteristics, as well as different device characteristics in different frequency ranges. Therefore, in the working frequency range, stable static working voltage and design biasing network are chosen on the basis of current–voltage (I-V) static curve of the power tube.

With the selection of the voltage for static working point finished, the next step is to design the biasing network of the circuit. The designed biasing network should allow the DC signal to reach the drain of the power tube smoothly from the DC source, while preventing the AC power leakage caused by insufficient input impedance (avoid the influence of AC biasing circuit on the impedance characteristics of the parts of AC circuit). This design adopts the passive biasing network, whose circuit is relatively simple, and usually composed of such passive devices as the inductor, the capacitor,

and the microstrip lines, providing RF power tube with proper working current and voltage.

The transistor used in this design is a kind of field effect tube, whose grid–source voltage is usually negative at the static working point, so the passive biasing network is adopted, and the bipolar power supply is needed. Two chokes (RFC) are utilized in the biasing circuit, which, respectively, connect the −2.9 V power supply and the grid, the 28 V power supply, and the drain. Meanwhile, some capacitance is connected in series to the devices for decoupling and filtering; bypass capacitors are connected in parallel to filter the fluctuation interference. Bypass capacitors establish for the power a channel of low-ground impedance, to filter the burr noises and reduce the transient current of the power supply module. In addition, the decoupling capacitor is added. Its main function is to remove interference of the signal of other frequency, but to stop the RF energy from going into the power distribution network, providing the circuit with the DC power supply.

3.3 Design of Input and Output Matching Network

The standard input and output impedance value of RF circuit is 50 Ω, but the input and output impedance of the power tube is not as big as that; therefore, before the design of input and output impedance matching circuit, the input and output impedance of the transistor must be measured first; the methods used are load pull and source pull.

Load pull is the determination of impedance value at the maximum output of power tube through the changing of load impedance value, which is guaranteed by the condition of matching input impedance. Similarly, the output impedance value of power tube can also be determined in the highest working efficiency. Meanwhile, load points of same power gain in Smith Chart are connected, and equal gain circle diagram is obtained, which can display very well the maximum achievable power output and efficiency of the circuit to be tested.

The control with variable load value is added to the circuit output, and the output impedance value is determined at the peak of output power and additional efficiency, according to the principle that transmitted power has to be set the highest.

4 Index Test for RF Power Amplifier

Serial input and output impedance matching circuits are, respectively, connected in serial to the grid and the drain of CGH40120F, with the static working voltage of the amplifier set to −2.9 V at the grid, and 28 V at the drain; the circuit is shown in Fig. 3; S parameter simulation and harmonic simulation are carried out on this circuit test. Such indicators as small signal gain, output power, input voltage standing wave ratio, and additional efficiency of the power amplifier are tested.

4.1 S Parameter Test

S parameter simulation tests two indicators of the power amplifier: the small signal gain and the input voltage standing wave ratio.

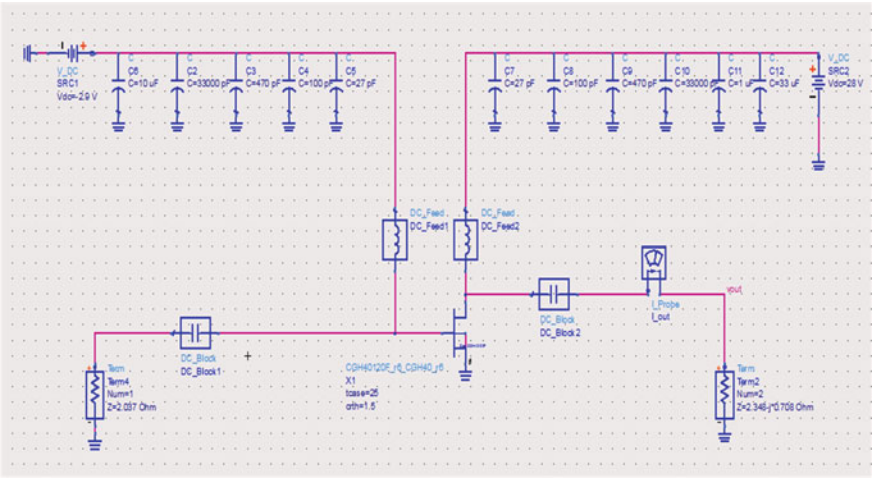


Fig. 3 Overall circuit of power amplifier

Figure 4 shows S (2, 1) curve graph of the power amplifier, namely, the small signal gain curve. It is concluded that throughout the working frequency range (950 MHz to 1250 MHz) the result is greater than 15 dB, which means indicators meet the requirements. Figure 5 shows the input voltage standing wave ratio of the power amplifier, which is less than 2:1 throughout the working frequency range, which means indicators meet the requirements.

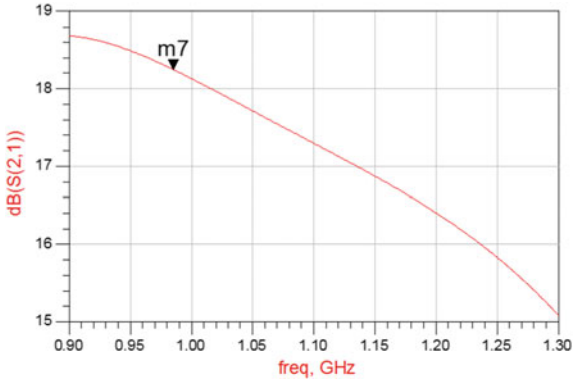


Fig. 4 Small signal gain of power amplifier

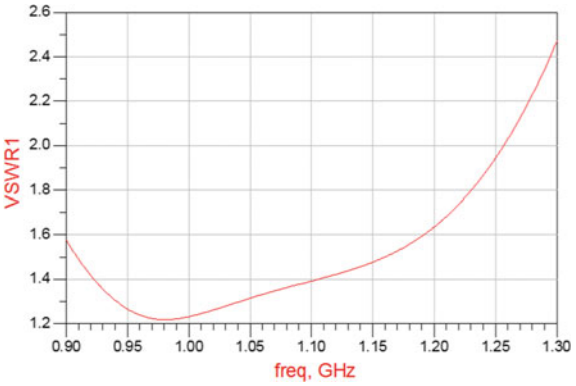


Fig. 5 Input voltage standing wave ratio of power amplifier

4.2 Harmonics Test

Harmonics simulation is to test the output power, harmonics suppression, and additional efficiency of the power amplifier.

(1) Test One

Test one is used to find out how every index changes with the input power under 1 Tone input signal, and the central working frequency of 1100 MHz;

In Fig. 6, when the working frequency is at 1100 MHz and the input power is 33 dBm, the linear increase no longer happens to the output power, so the 1 db compression point of the amplifier is 37 dBm, with the output power of the moment being 51.4 dBm.

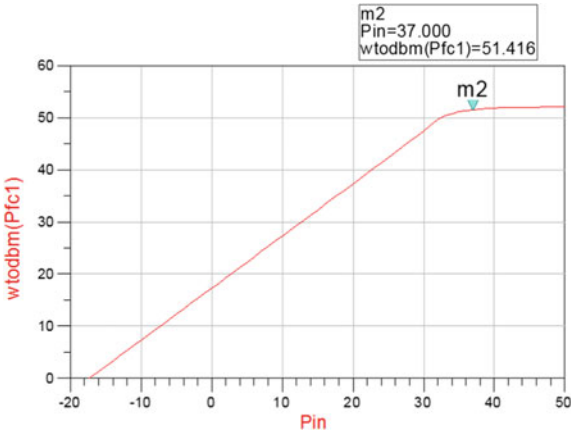


Fig. 6 At frequency of 1100 MHz, output power changes with input power

(2) Test Two

Test two is used to find out how each index changes with the frequency under 1 Tone input signal, and the input power of 34 dBm.

Figure 7 shows under the input power of 34 dBm, the output power fluctuations near 50.1 dBm (100 W) throughout the working frequency range (950 MHz to 1250 MHz), with the flatness being ±0.1 dB. Figure 8 shows under the input power of 34 dBm, and the added efficiency reaches the bottom (equals to 47%) at the working frequency range of 950 MHz within the whole spectrum. The indicators are in line with the requirements that added efficiency is more than 45%.

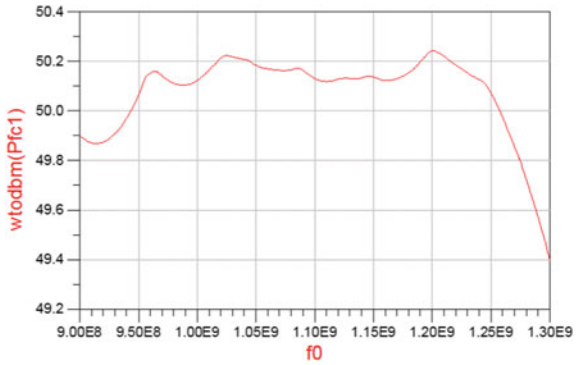


Fig. 7 At the input power of 34 dBm, output power changes with frequency

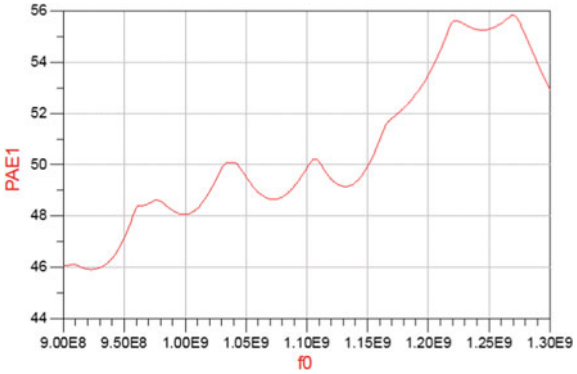


Fig. 8 At the input power of 34 dBm, added efficiency changes with frequency

5 Summarization and Prospect

In this paper, CGH40120F GaN high-electron-mobility transistor (HEMT) of Cree Company is utilized for the design of L-band high power amplifier, and its simulation is carried out on the ADS platform. When RF power amplifier is designed, DC test is implemented on the power tube first, to determine its static working point; then the biasing circuit and stability circuit are designed to ensure that power tube stays in absolute stability within the whole spectrum, so as to avoid the serious consequences caused by self-oscillation.

The current broadband RF amplifier is still a problem to overcome, so this design is to be improved in many aspects. On one hand, due to the broad working bandwidth, it is difficult to achieve the coherent linearity in the full frequency range; if the linearity demand is high, extra circuits are needed to improve the linearity. Therefore, the abovementioned aspects can be explored in further studies and researches.

References

1. Sun, K., Wu, D.: MPC-based delay-aware fountain codes for live video streaming. In: IEEE International Conference on Communications, ICC (2016)
2. Schmidt, R.O.: Multiple emitter location and signal parameter estimation. IEEE Trans. Antennas Propag. **34**(3), 276–280 (1986)
3. Sun, K., Zhang, H., Wu, D., Zhuang, H.: MPC-based delay-aware fountain codes for real-time video communication. IEEE Internet Things J. (IoT) (2016)
4. Tayem, N., Kwon, H.M.: L-shape 2-dimensional arrival angle estimation with propagator method. IEEE Trans. Antennas Propag. **53**(5), 1622–1630 (2005)
5. Liu X, Li, Z.: Information-centric mobile ad hoc networks and content routing: a survey. Article in press (2016)
6. Gruteser, M., Grunwald, D.: Anonymous usage of location based services through spatial and temporal cloaking. In: ACM/USENIX MobiSys (2003)
7. Gu, J.F., Zhu, W.P., Swamy, M.N.S.: Joint 2-D DOA estimation via sparse L-shaped array. IEEE Trans. Signal Process. **31**(5), 1171–1182 (2015)
8. Sun, K., Wu, D.: Video rate control strategies for cloud gaming. J. Vis. Commun. Image Represent. (2013)
9. Li, Z., Chen, Y.T., Shi, H., Liu, K.: NDN-GSM-R: a novel high-speed railway communication system via named data networking. EURASIP J. Wirel. Commun. Netw. (2016)
10. Sun, K., Yan, B., Gharavi, H.: Low complexity content-aware image retar-geting. In: IEEE International Conference on Image Processing, ICIP 2012 (2012)
11. Ghinita, G., Kalnis, P., Skiadopoulos, S.: PRIVE: anonymous location based queries in distributed mobile systems. In: Proceedings of International Conference on World Wide Web (WWW 2007), Banff, Alberta, Canada, pp. 1–10 (2007)
12. Sun, K., Wu, D.: Video rate control strategies for cloud gaming. J. Vis. Commun. Image Represent. (2013)
13. Li, Z., Liu, K.: MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11) (2014)
14. Xia, T.Q., Zheng, Y., Wan, Q., Wang, X.G.: 2-D angle of arrival estimation with two parallel uniform linear arrays for coherent signals. In: IEEE Radar Conference, vol. 55, no. 9, pp. 244–247(2007)
15. Li, Z., Song, L., Shi, H.: Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme. Article in press (2016)
16. Ming, T., Wu, Q., Guoping, Z., Lili, H.: A New Scheme of LBS Privacy Protection. IEEE

An Improved Codebook Design for 3D-MIMO Multiuser Precoding Scheme

Yinghong Cao[1], Sheng Wang[1], Peng Li[1], and Jue Wang[2(✉)]

[1] School of Information Science and Engineering, Dalian Polytechnic University, Dalian, China
{caoyinghong,lipeng}@dlpu.edu.cn, 18340854717@163.com
[2] Dalian Polytechnic University, 116034 Dalian, China
eecyh@hotmail.com

Abstract. In this paper, we concentrate on the codebook design for 3D-MIMO multiuser precoding scheme. An improved 3D precoding codebook design is proposed by considering the relative phase between the horizontal codebook and the vertical codebook. By setting the initial phase of the horizontal codebook properly, the distance between codewords can be maximized, so that the beam direction corresponding to the codewords can be more accurately pointed to the expected user in the real 3D space, and the spatial information of the horizontal and vertical dimensions can be more fully utilized. Simulation results demonstrate that, compared to the traditional 3D codebook design, this improved codebook design can suppress the multiuser co-channel interference more efficiently and improve the system performance effectively.

Keywords: Codebook design · 3D-MIMO · Multiuser · Co-channel interference · Massive MIMO

1 Introduction

Multiuser precoding technology provides substantial gains in co-channel interference mitigation by exploiting the spatial degrees of freedom. As a key technology for next-generation wireless communication system, three-dimensional (3D) MIMO multiuser precoding has attracted considerable interest in recent years [1]. With large antenna arrays equipped at the base station (BS), 3D-MIMO multiuser precoding schemes can form the transmitting beams point to the expected users more accurately. Accordingly, the spatial information of the horizontal and vertical dimensions can be more fully utilized to explore the potential of wireless communication system in real 3D space [2–6].

Codebook design for 3D-MIMO multiuser precoding schemes is one of the important issues that have been studied widely [7–12]. Each codeword in the codebook represents a precoding matrix, which also determines the direction of the transmitting beam. Therefore, to suppress the multiuser co-channel interference efficiently, the distance between codewords was expected to be maximal.

3D codebook design usually focuses on the horizontal direction codebook design, the vertical codebook design, and the final 3D codebook construction [10–12].

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 57–64, 2019.
https://doi.org/10.1007/978-981-13-6264-4_7

According to the construction of 3D codeword, prior works mostly ignore the relative phase problem between the horizontal codebook and the vertical codebook [10–12], which makes the inter-codeword distance relatively close. This causes the overlapping shadow part of the main lobe beams to be overlapped too much. And the interference suppression between multiusers still needs to be further improved.

In this paper, the codebook design for 3D-MIMO multiuser precoding is considered. An improved Kronecker-product-based codebook design is proposed. In this proposed codebook design, the relative phase between the horizontal codebook and vertical codebook was considered during the construction of 3D codeword. By setting the initial phase of the horizontal codebook properly, the distance between codewords can be maximized, so that the beam direction corresponding to the codewords can be more accurately pointed to the expected user in the real 3D space, and the spatial information of the horizontal and vertical dimensions can be more fully utilized. Simulation results demonstrate the performance of our scheme.

2 System Model

Consider a single-cell 3D-MIMO multiuser downlink system; the BS is equipped with $N \times M$ transmit antennas, where N and M represent the number of horizontal and vertical antenna elements of the uniform planar array (UPA), respectively. Assume that there are N_u co-channel users distributed in the cell, and each of the users is equipped with one receive antenna.

We consider a 3D channel model, and the channel is assumed to consist of L equal gain NLOS paths. For the lth path, let φ_l emulate the random phase shift from different lengths of the transmit paths, which is uniformly distributed in $[0, 2\pi]$. Then, the channel response can be formulated as [9]

$$\boldsymbol{h} = \sum\nolimits_{l=1}^{L} \frac{e^{j\varphi_l}}{\sqrt{L}} \left[1, e^{-ju_l}, \cdots, e^{-j(M-1)u_l}\right]^T \left[1, e^{-jv_l}, \cdots, e^{-j(N-1)v_l}\right] \tag{1}$$

where $u_l = \frac{2\pi d_v}{\lambda} \sin(\theta_l + \Delta\theta_l) \cos(\phi_l + \Delta\phi_l)$, $v_l = \frac{2\pi d_h}{\lambda} \sin(\theta_l + \Delta\theta_l) \sin(\phi_l + \Delta\phi_l)$. In the above formula, λ is the carrier wavelength; d_h and d_v represent the space between antenna elements in horizontal and vertical, respectively; ϕ_l is the transmit angle-of-departure (AoD) in azimuth, $\Delta\phi_l$ is the perturbation in horizontal, which is normal distributed as $N(0,\ \sigma)$, and σ is the standard deviation. Similarly, θ_l is the transmit AoD in elevation, $\Delta\theta_l$ is the perturbation in vertical, which is assumed to be normal distributed as $N(0,\ \xi)$, and ξ is the standard deviation.

Let s_i denote the transmitted data for user $i (i = 1, \cdots, N_u)$, and satisfy $E[s_i^* s_i] = 1$. The received signal at the ith user can be expressed as

$$\boldsymbol{y}_i = \sqrt{\frac{P}{N_u}} \boldsymbol{h}_i \boldsymbol{w}_i s_i + \sqrt{\frac{P}{N_u}} \sum\nolimits_{j=1, j\neq i}^{N_u} \boldsymbol{h}_i \boldsymbol{w}_j s_j + \boldsymbol{n}_i \tag{2}$$

where P is the total transmit power, and P/N_u is the transmit power of each user based on equal power allocation; $\boldsymbol{h}_i \in \mathrm{C}^{1\times M\times N}$ denotes the 3D channel state information

matrix from BS to user i; $\boldsymbol{w}_i \in \mathrm{C}^{M\times N\times 1}$ is the precoding matrix for user i; $\boldsymbol{n}_i \sim \boldsymbol{CN}(0, \sigma_n^2)$ is the complex additive white Gaussian noise.

In downlink transmission, for the codebook-based precoding scheme, precoding matrix index (PMI) is selected at the user receiver according to the estimated instantaneous channel state information (ICSI). Then, the index information is fed back to the BS through a feedback link. Based on these feedback information, the BS selects appropriate 3D codewords to mitigate multiuser interference.

3 Proposed Codebook Design for 3D-MIMO Precoding

In the codebook for 3D-MIMO multiuser precoding, each codeword represents a precoding matrix, which also determines different directions of the transmitting beam. In order to reduce co-channel interference, the distance between codewords was expected to be maximal. In this section, we illustrate the way of constructing our proposed codebook. Assume that a number of codewords in the vertical and horizontal dimensions are N_v and N_h, respectively. Then, in 3D space, the beamforming coverage area is divided into N_v vertical ring areas, and each ring area is divided into N_h beam directions (shown in Fig. 1). The steps of generating the proposed codebook are as follows:

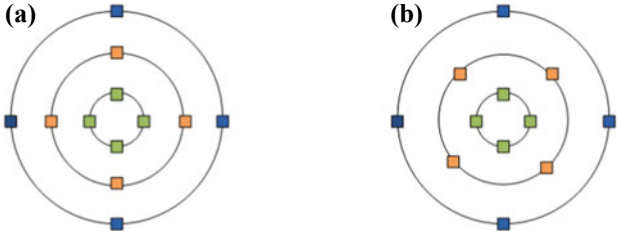


Fig. 1. Illustration of beam centers according to the codewords ($N_v = N_h = 4$). (a) Conventional codebook [10] (b) Proposed codebook

Horizontal codebook design Traditional Discrete Fourier Transform (DFT) codebook is suitable for horizontal beamforming, since the beams corresponding to the DFT codewords are approximately equally distributed on a circle, which provides a good fit to the horizontal beam characteristic [10]. Therefore, based on the DFT codebook, the nth $(n = 0, \cdots, N_h - 1)$ horizontal codeword can be given as

$$\boldsymbol{f}_h^n = \frac{1}{\sqrt{N_h}}\left[1, e^{j\frac{2\pi n}{N_h}}, \cdots, e^{j\frac{2\pi(N-1)n}{N_h}}\right] \tag{3}$$

Vertical codebook design Due to the limited range of downtilt, DFT codebook is no longer suitable for the vertical beamforming. A spatial-segmentation-based vertical codeword design is introduced [12].

As shown in Fig. 1, N_v vertical codewords divide the cell into N_v equal distance ring areas. Each vertical beamforming points to the center of the corresponding ring.

Let H_{BS} denote the height of the base station antenna, H_{MS} denote the height of user antenna, and S_m denote the horizontal distance from the mth ring center $(m = 0, \cdots, N_v - 1)$ to the base station. Then, with respect to the base station antenna, for the mth vertical codeword, the downtilt angle of the mth ring center θ_m can be expressed as

$$\theta_m = \arctan\left(\frac{S_m}{H_{BS} - H_{MS}}\right) \tag{4}$$

Since the phase difference for the transmitter antenna array is $\frac{2\pi d_v \cos\theta_m}{\lambda}$, the mth vertical codeword can be given as

$$\boldsymbol{f}_v^m = \frac{1}{\sqrt{N_v}}\left[1, e^{j\frac{2\pi d_v \cos\theta_m}{\lambda}}, \cdots, e^{j\frac{2\pi d_v (N_v-1)\cos\theta_m}{\lambda}}\right]^T \tag{5}$$

3D codebook construction The final codeword is the Kronecker product of horizontal and vertical codeword. During the construction of 3D codeword, the relative phase between the horizontal codebook and vertical codebook was considered in this proposed codebook design, which is different from the conventional Kronecker-product-based codebook design [9–12], i.e., the proposed codeword $\boldsymbol{c}_{m,n}^{prop}$ is given by

$$\boldsymbol{c}_{m,n}^{prop} = e^{j\omega_m}\left(\boldsymbol{f}_v^m \otimes \boldsymbol{f}_h^n\right) \quad \text{where } \omega_m = \begin{cases} 0, & m \text{ is odd} \\ \frac{\pi}{N_h}, & m \text{ is even} \end{cases} \tag{6}$$

Here, $m = 0, \cdots, N_v - 1$; $n = 0, \cdots, N_h - 1$; $\otimes$ denotes the Kronecker product; and ω_m represents the horizontal phase rotation angle corresponding to the codewords that belongs to the mth ring area.

The comparison of the beam center distribution, between conventional codebook and the proposed codebook, is shown in Fig. 1. The color squares represent the beam directions according to different codewords. According to their position, users can select the preferred codewords as precoding matrix. As shown in Fig. 1b, the color squares on two adjacent rings have different initial phases. By setting the initial phase properly, the distance between codewords can be maximized. As a result, the overlapping area of different beams is smaller than the traditional codebook, and the inter-user interference can be reduced.

4 Codewords' Selection

In the codebook-based 3D-MIMO multiuser precoding scheme, the user selects the optimal precoding codeword based on the maximum received power criterion, that is, the optimal codeword index for user $i\,(i = 1, \cdots, N_u)$ is selected by

$$\mathrm{PMI}_i^{opt} = \underset{m=0,\cdots,N_v-1;\ n=0,\cdots,N_h-1}{\arg\max} \left\| \boldsymbol{h}_i \boldsymbol{c}_{m,n}^{prop} \right\|_F^2 \tag{7}$$

Then, the index information is fed back to the BS through a feedback link. And based on the feedback $\{\mathrm{PIM}_i\}_{i=1}^{N_u}$, the BS selects appropriate 3D codeword as precoding matrix to mitigate multiuser interference.

5 Simulation Results

Simulation results are presented to demonstrate the performance of the proposed codebook design.

Consider a single-cell multiuser wireless communication system, assuming a uniform spatial user distribution over the cell, and the cell radius is $R = 500$ m. The base station (BS) is located in the center of the cell and is equipped with $N \times M$ UPA antenna, with the antenna element spacing $d_v = d_h = \lambda/2$. The BS antenna's height is $H_{BS} = 50$ m. Assume each of the users is equipped with a single receive antenna, and the user antenna's height is $H_{MS} = 1$ m.

For the wireless channel, there are $L = 20$ equal gain NLOS paths; the transmit AoD in azimuth is $\phi_1 = \pi/3$, the standard deviation for perturbation in horizontal is $\sigma = \pi/12$, the transmit AoD in elevation is $\theta_1 = 3\pi/18$, and the standard deviation for perturbation in vertical is $\xi = \pi/36$.

The BER and CDF performances of the 3D-MIMO multiuser precoding systems with the proposed codebook are presented with different numbers of transmit antennas and co-channel users. For performance comparison, the performances of the 3D-MIMO multiuser precoding systems with the DFT codebook [10] and the conventional codebook [12] are presented, respectively.

Figures 2 and 3 demonstrate the difference in BER performance among the 3D-MIMO multiuser precoding system with the DFT codebook [10], the conventional codebook [12], and the proposed codebook. In Fig. 2, the transmit antenna is $N \times M = 8 \times 8$, the number of co-channel users is $N_u = 8$, and the size of codebook is $N_v \times N_h = 8 \times 8$. In Fig. 3, the transmit antenna is $N \times M = 8 \times 4$, the number of co-channel users is $N_u = 4$, and the size of codebook is $N_v \times N_h = 4 \times 8$. From Figs. 2 and 3, it can be seen that, under the same system configure, the BER performance of the systems with proposed codebook outperforms the other two systems. With the increase of SNR, the BER performance of the proposed system improves more significantly. Therefore, based on this proposed codebook design, the 3D-MIMO multiuser precoding system can suppress the multiuser co-channel interference more efficiently (Figs. 2 and 3).

Figures 4 and 5 provide the throughput performance curve among the 3D-MIMO multiuser precoding system with the DFT codebook [10], the conventional codebook [12], and the proposed codebook. In Fig. 4, the transmit antenna is $N \times M = 8 \times 8$, the number of co-channel users is $N_u = 8$, and the size of codebook is $N_v \times N_h = 8 \times 8$. In Fig. 5, the transmit antenna is $N \times M = 8 \times 4$, the number of co-channel users is $N_u = 4$, and the size of codebook is $N_v \times N_h = 4 \times 8$. From Figs. 4 and 5, it is obvious

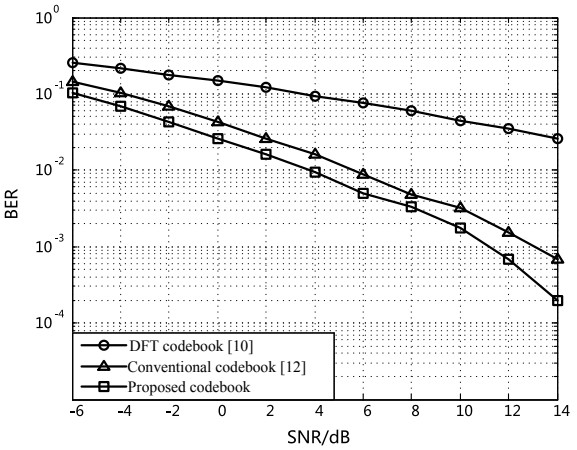


Fig. 2. BER performance of different codebook designs ($M = 8$, $N = 8$, $N_u = 8$)

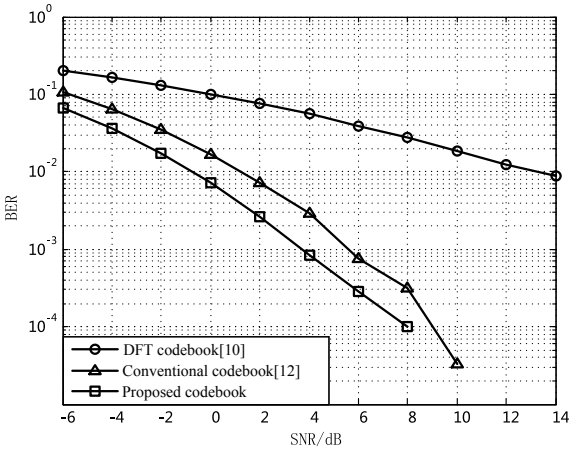


Fig. 3. BER performance of different codebook designs ($M = 4$, $N = 8$, $N_u = 8$)

that, under the same system configure, the throughput performance of the proposed system is better than the other two systems. For the same CDF value, the system with the proposed codebook can obtain more throughput gain. Hence, the 3D-MIMO multiuser precoding system based on the proposed codebook can make better use of the spatial dimension and achieve the improvement of the system effectiveness.

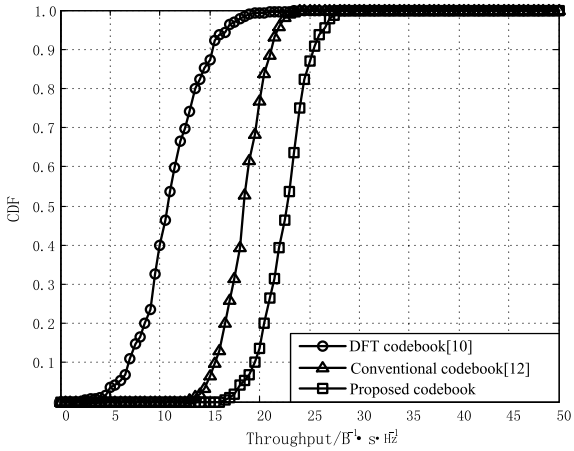


Fig. 4. CDF of different codebook designs ($M = 8$, $N = 8$, $N_u = 8$)

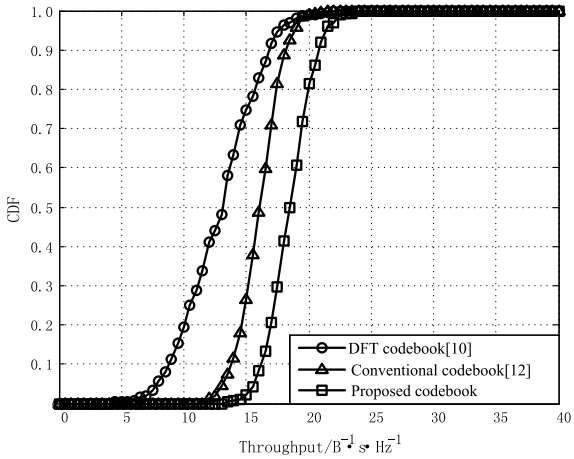


Fig. 5. CDF of different codebook designs ($M = 4$, $N = 8$, $N_u = 8$)

6 Conclusion

An improved codebook design is proposed for 3D-MIMO multiuser precoding scheme. By considering the relative phase between horizontal codebook and vertical codebook during the construction of codewords, this proposed 3D codebook design can achieve a maximum distance between codewords; thereby, the beam directions corresponding to the codeword can be more accurately pointed to the desired user, and the co-channel interference can be mitigated more effectively. Simulations demonstrate the performance of the proposed codebook design.

Acknowledgements. This work was supported by Liaoning Provincial Department of Education Research Project (L2015043) and Doctoral Scientific Research Foundation of Liaoning Province (201601280, 201501191).

References

1. Nam, Y.H., Ng, B.L., Sayana, K., et al.: Full-dimension MIMO (FD-MIMO) for next generation cellular technology. IEEE Commun. Mag. **51**, 172–179 (2013)
2. Hui, T., Fei, W., Xueying, H., et al.: A novel limited feedback mechanism based on mixed reciprocity and codebook in large scale 3D-MIMO systems. IEEE Signal Inf. Process, 176–180 (2016)
3. 3GPP: Precoding schemes for elevation beamforming and FU-MIMO. TSG RAN1, R1-144976 (2014)
4. Love, D.J., Heath, R.W., Strohmer T.: Grassmannian beamforming for multiple-input multiple-output wireless systems. IEEE Trans. Inf. Theory **49**, 2735–2747 (2003)
5. Wentian, Z., Wei, X., Angyan, L., et al.: 3D MIMO Kronecker channel modeling based on three typical antenna arrays. Syst. Eng. Electron. **39**, 1366–1373 (2017)
6. Penna, F., Hongbing, C., Jungwon, L.: Low complexity precoder selection for FD-MIMO systems. In: Vehicular Technology Conference, IEEE, pp. 1–5 (2017)
7. Qiao, D., Qian, H., Li, G.Y.: Multi-resolution codebook design for two-stage precoding in FDD massive MIMO networks. In: International Workshop on Signal Processing Advances in Wireless Communications, IEEE, pp. 1–5 (2017)
8. Fangchao, Z., Shaohui, S., et al.: Hybrid CSI-RS transmission mechanism-based 3D beamforming scheme for FDD massive MIMO system. China Commun. **13**(2), 109–119 (2017)
9. Ying, D., Vook, F.W., Thomas, T.A., et al.: Kronecker product correlation model and limited feedback codebook design in a 3D channel model, pp. 5865–5870 (2014)
10. Yi, X., Shi, J., Jue, W., et al.: A limited feedback scheme for 3D multiuser MIMO based on Kronecker product codebook. In: International Symposium on Personal Indoor and Mobile Radio Communications, IEEE, pp. 1130–1135 (2013)
11. Youcai, Z., Xiaorong, J.: Multi-user precoding scheme based on 3D codebook in massive MIMO system. Comput. Eng. **43**, 69–74 (2017)
12. Fan, Z., Shiwen, H., Yongming, H., et al.: Limited feedback based massive 3D MIMO multi-user transmission method and its codebook design. J. Data Acquis. Process. **29**, 384–389 (2014)

A New Nonlinear Precoding Scheme for Spatial Modulation

Jinwei Huang, Wenbin Zhang(✉), Yingze Xie, and Shaochuan Wu

Communication Research Center, Harbin Institute of Technology, Harbin, China
zwbgxy1973@hit.edu.cn

Abstract. Considering most previous literature focus on the linear scheme in precoding spatial modulation, this paper proposes a novel nonlinear algorithm for multiuser generalized spatial modulation, which is based on Tomlinson–Harashima precoding (THP). We call it reordered generalized precoding spatial modulation (RGPSM). This algorithm introduces the "best-first" method to rearrange the index of mobiles in precoding procedure. In addition, this algorithm can provide multi-stream for each user. The simulation results show that our algorithm is better than unordered algorithm in BER.

Keywords: MIMO · GPSM · Tomlinson–Harashima precoding (THP) · Ordering

1 Introduction

In order to achieve higher throughput, classical precoding spatial modulation (PSM) scheme can be extended to a multi-stream case [1]. In this case, more active streams are simultaneously transmitted and the transmission scheme is called generalized precoding-aided spatial modulation (GPSM). Analytical studies of GPSM are revealed in [2], which provides the upper bound of both the symbol error ratio (SER) and bit error ratio (BER) expressions of the GPSM scheme. Inspired by the work of [1], many papers about GPSM have been published. In [3], a scheme named MSR-SM is proposed, which proposes a multi-stream spatial modulation at the receiver, where both accurate BER expressions and the achievable diversity gains are revealed. In [4], the conventional SM and the GPSM schemes were beneficially combined into hybrid SM for dual-hop relay-aided communications. [5] proposes GVSM and CI-GVSM schemes that apply SM principle to the virtual parallel channels resulted from the SVD of the MIMO channels. Another variant of PSM called generalized precoding-aided quadrature SM (GPQSM) is proposed in [6], where each complex M-ary symbol is separated into real and imaginary part, and uses the SM principle individually.

This work was supported by National Natural Science Foundation (NNFS) of China under Grant 61671173.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 65–72, 2019.
https://doi.org/10.1007/978-981-13-6264-4_8

In contrast to most existing GPSM schemes, relying on linear precoding [6], it appears beneficial to explore the nonlinear GPSM scheme, and some papers have been proposed on the nonlinear schemes. However, these schemes assume that each user has a single stream in multiuser communication system. So, this paper proposes a novel algorithm, which extends a single-stream to a multi-stream system, and this algorithm introduces a sorting method into a nonlinear precoding technique, Tomlinson–Harashima precoding (THP). We call this algorithm reordered generalized precoding spatial modulation (RGPSM), to the best of our knowledge, no work regarding the ordering of GPSM is available in the literature.

The rest of the paper is organized as follows. In Sect. 2, we introduce the GPSM MIMO system model. This is followed by THP scheme analysis in Sect. 3. The ordering problem of this algorithm is addressed in Sect. 4. Our simulation results are provided in Sect. 5, and finally we conclude in Sect. 6.

Notations: Boldface uppercase, boldface lowercase, and lowercase letter $\mathbf{A}$, $\mathbf{a}$, and a denote a matrix, vector, and scalar variables, respectively. $\mathbf{A}^T$, $\mathbf{A}^H$, and $\mathbf{A}^{-1}$ represent the transpose, the conjugate transpose, and the inverse of a matrix, respectively. $E()$ denotes expectation, $||$ denotes absolute value, $Tr()$ denotes trace of a matrix, and *diag* denotes diagonal.

2 System Model

Reference [7] discusses the MU-MIMO downlink system with single-stream and single-user antenna. We extend it to a multi-stream and multiuser antennas system as shown in Fig. 1. This system includes a base station with N_t antennas and U users. ith user has N_i antennas, and the sum of user antennas is $N_r = \sum_{i=1}^{U} N_i$, the number of data streams for each user is the same as its antenna number, and N_a represents numbers of activated antennas in spatial modulation. Data streams for each user are denoted as $\mathbf{a}_i^{'}(n) = [a_{i1}^{'}(n), a_{i1}^{'}(n), \ldots, a_{iN_i}^{'}(n)], i \in [1, 2, \ldots, U]$. Denoting $\mathbf{P}_i$ and $\mathbf{W}_i, i \in [1, 2, \ldots, U]$ as spatial modulation and MU-MIMO precoding matrix for each user, respectively. In this paper, it is the spatial modulation matrix $\mathbf{P}_i$ that ensures each antenna of each user receives one data stream. Passing precoding matrix $\mathbf{W}_i$, $\mathbf{a}_i^{'}(n)$ is translated into $\mathbf{W}_i\mathbf{P}_i\mathbf{a}_i^{'}(n)$. So, the signal transmitted by base station is written as

$$\mathbf{x}(n) = \sum_{i=1}^{U} \mathbf{W}_i\mathbf{P}_i\mathbf{a}_i^{'}(n) \,, \tag{1}$$

where $\mathbf{a}_i^{'}(n), \mathbf{P}_i, \mathbf{W}_i$ must satisfy the normalization conditions of [8], which means

$$E[\mathbf{a}_i^{'}(n)\mathbf{a}_i^{'}(n)]^H = \frac{1}{N_i}\mathbf{I} \,. \tag{2}$$

$$Tr(\mathbf{W}_i\mathbf{P}_i\mathbf{P}_i^H\mathbf{W}_i^H) = N_i \,. \tag{3}$$

After transmission over MIMO channel, the received signal by each user can be written as

$$\mathbf{y}_i(n) = \mathbf{H}_i \mathbf{W}_i \mathbf{P}_i \mathbf{a}_i^{'}(n) + \mathbf{H}_i \sum_{u \neq i}^{U} \mathbf{W}_u \mathbf{P}_u \mathbf{a}_u^{'}(n) + \mathbf{n}_i \quad i \in [1, 2, \ldots, U] \,. \tag{4}$$

For convenience, we ignore the time index n in (4), and rewrite (4) as

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{W}_i \mathbf{P}_i \mathbf{a}_i^{'} + \mathbf{H}_i \sum_{u \neq i}^{U} \mathbf{W}_u \mathbf{P}_u \mathbf{a}_u^{'} + \mathbf{n}_i \quad i \in [1, 2, \ldots, U] \,. \tag{5}$$

We introduce $\mathbf{a}_i(n) = \mathbf{P}_i \mathbf{a}_i^{'}$, and Eq. (5) can be expressed as

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{W}_i \mathbf{a}_i + \mathbf{H}_i \sum_{u \neq i}^{U} \mathbf{W}_u \mathbf{a}_u + \mathbf{n}_i \quad i \in [1, 2, \ldots, U] \,, \tag{6}$$

where $\mathbf{H}_i$ is the channel matrix between the user i and the base station, whose elements follow the independent and identical cyclic complex Gaussian distribution with $CN(0, 1)$, and $\mathbf{n}_i$ is the Gaussian noise vector with $(0, \sigma_i^2 \mathbf{I})$.

We denote $[\mathbf{H}_1^T, \mathbf{H}_2^T, \ldots, \mathbf{H}_U^T]^T$ as $\mathbf{H}$, which is a $\sum_{i=1}^{U} N_i$ by N_t matrix. Considering $E[Tr(\mathbf{H}\mathbf{H}^H)]$ equals $N_r \times N_t$, and $E[\mathbf{H}\mathbf{H}^H] = \mathbf{I}$, by (2) and (3), the power of received signal is normalized at each receive antenna. So the SNR of each receive antenna equals $1/\sigma_i^2$.

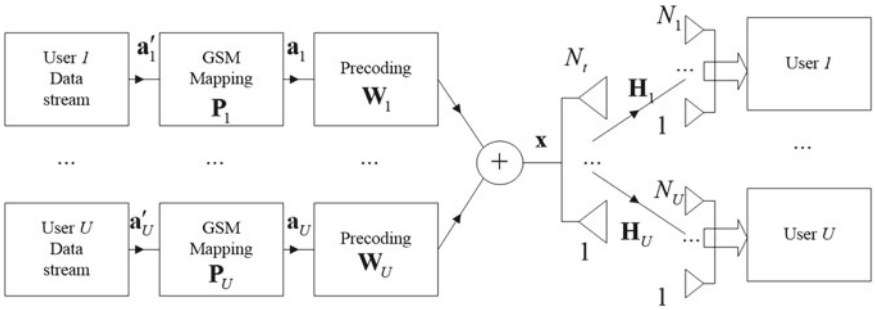


Fig. 1. MU-MIMO GPSM transmission system model.

3 THP Multiuser Spatial Modulation Precoding Scheme

Substituting $\mathbf{W}_i, i \in [1, 2, \ldots, U]$ with THP, we obtain the THP structure as shown in Fig. 2. In this figure, $N_r = N_t = N$, $\mathbf{F}$ stands for feedforward matrix,

MOD stands for the calculation of module $2\sqrt{M}$, where M means $M-QAM$, and **B-I** stands for unit lower triangular matrix. Thus, $\tilde{\mathbf{a}}_i$ can be written as

$$\tilde{\mathbf{a}}_i = mod_{2\sqrt{M}}(\mathbf{a}_i - (\mathbf{B\text{-}I})\tilde{\mathbf{a}}_i) = \mathbf{a}_i - (\mathbf{B\text{-}I})\tilde{\mathbf{a}}_i + \mathbf{d}\ , \tag{7}$$

where $\mathbf{d}$ is the precoding vector used to constrain the value of $\tilde{\mathbf{a}}_i$. Let $\mathbf{v} = \mathbf{a}_i + \mathbf{d}$, (7) can be rewritten as

$$\tilde{\mathbf{a}}_i = \mathbf{B}^{-1}(\mathbf{a}_i + \mathbf{d}) = \mathbf{B}^{-1}\mathbf{v}\ . \tag{8}$$

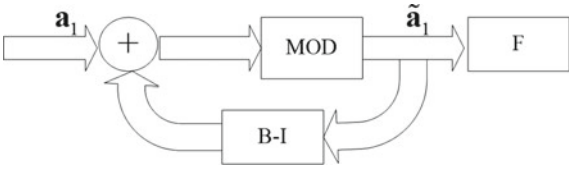


Fig. 2. Equivalent precoding model.

It is worth to note that THP is a nonlinear coding, whose output is not constrained to the constellation points. This leads to more power consumed at transmitter according to reference [7].

After passing the MIMO channel, the received signal is written as

$$\mathbf{r} = \mathbf{HFB}^{-1}\mathbf{v} + \mathbf{n}\ , \tag{9}$$

where $\mathbf{r} = [\mathbf{r}_1^T, \mathbf{r}_2^T, \ldots, \mathbf{r}_U^T]^T$, $\mathbf{n} = [\mathbf{n}_1^T, \mathbf{n}_2^T, \ldots, \mathbf{n}_U^T]^T$, $\mathbf{H} = [\mathbf{H}_1^T, \mathbf{H}_2^T, \ldots, \mathbf{H}_U^T]^T$, and $\mathbf{R}_n = E[\mathbf{nn}^H] = \sigma_n\mathbf{I}$.

At the receiver, introduce a weighting matrix $\mathbf{G} = diag[\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_U]$, whose diagonal elements are the coefficient factors of each user receiver. And we can derive

$$\mathbf{y}' = \mathbf{Gr} = \mathbf{GHFB}^{-1}\mathbf{v} + \mathbf{Gn} = \mathbf{GHFB}^{-1}\mathbf{v} + \tilde{\mathbf{n}}\ , \tag{10}$$

where $\mathbf{y}' = [\mathbf{y}_1'^T, \mathbf{y}_2'^T, \ldots, \mathbf{y}_U'^T]^T$, and $\tilde{\mathbf{n}} = \mathbf{Gn} = [\tilde{\mathbf{n}}_1'^T, \tilde{\mathbf{n}}'^T, \ldots, \tilde{\mathbf{n}}_U'^T]^T$. If zero-forcing criterion is used, then

$$\mathbf{GHFB}^{-1} = \mathbf{I} \tag{11}$$

is required. Decomposing $\mathbf{H}^H$ into $\mathbf{QR}$ by QR factorization, $\mathbf{R} = [r_{ij}]$ is an upper triangular matrix with real diagonal elements. Next, we can calculate the process matrix $\mathbf{B}$, the feedforward matrix $\mathbf{F}$, and the weighting matrix $\mathbf{G}$:

$$\mathbf{G} = diag[r_{11}^{-1}, r_{22}^{-1}, \ldots, r_{NN}^{-1}], \quad \mathbf{B} = \mathbf{GR}^H, \quad \mathbf{F} = \mathbf{Q}\ . \tag{12}$$

By (12), $\mathbf{y}'$ can be rewritten as

$$\mathbf{y}' = \mathbf{Q} + \tilde{\mathbf{n}}\ . \tag{13}$$

Observing (13), we can see that all the interstream interference has been eliminated. The detection of each user depends on $\tilde{\mathbf{n}}$, whose covariance matrix is written as

$$\Phi_{\tilde{n}\tilde{n}} = E[\tilde{\mathbf{n}}\tilde{\mathbf{n}}^H] = diag[\sigma_n^2/r_{11}, \sigma_n^2/r_{22}, \ldots, \sigma_n^2/r_{NN}] \,. \tag{14}$$

4 A New Algorithm in the THP-GPSM System

Each user is different in BER for THP-GPSM, and the worst one has the largest impact on the whole system [9]. We propose a novel nonlinear algorithm reordered generalized precoding spatial modulation (RGPSM) to tackle with this problem. This algorithm precodes its data stream in descending order by SNR. The details of algorithm are shown in Algorithm 1, and the symbols and formulas involved in this algorithm are explained as follows.

$V = \{v_1, v_2, \ldots, v_U\}$ is composed of some unsorted users, and we will rearrange the position each user in set V according to SNR. For user v_i, we can calculate the noise variance by $\sum_{m=1}^{N_i} E[|\tilde{n}_{v_{jm}}|^2]$, where subscript N_i denotes antenna number for user v_j. This formula is the sum of every antenna SNR of user v_j. The smaller it is, the greater the SNR of the user v_j is. For user v_j, referring to [7], we have $\sum_{m=1}^{N_j} E[|\tilde{n}_{v_{jm}}|^2] = sum\{\sigma_n^2(diag(\mathbf{H}_{v_j}\mathbf{H}_{v_j}^H))\}$, $\mathbf{H} = [\mathbf{H}_{v_1}^T, \mathbf{H}_{v_2}^T, \ldots, \mathbf{H}_{v_U}^T]^T$. Calculating the SNR for each user, we can obtain the ordered set $L = \{l_1, l_2, \ldots, l_U\}$ where l_1 and l_U have the highest and the lowest SNR, respectively. Following L, rewrite the channel matrix as

$$\mathbf{H}_L = [\mathbf{H}_{l_U}^T, \mathbf{H}_{l_{U-1}}^T, \ldots, \mathbf{H}_{l_1}^T]^T \,. \tag{15}$$

The data vector is also rearranged as $\mathbf{a}_L = [\mathbf{a}_{l_U}^T, \mathbf{a}_{l_{U-1}}^T, \ldots, \mathbf{a}_{l_1}^T]^T$, and received vector is $\mathbf{y}_L = [\mathbf{y}_{l_U}^T, \mathbf{y}_{l_{U-1}}^T, \ldots, \mathbf{y}_{l_1}^T]^T$. By this new channel matrix, the corresponding processing matrix $\mathbf{B}_L$, $\mathbf{F}_L$, and $\mathbf{G}_L$ can be generated according to the procedure in Sect. 3.

Algorithm 1. RGPSM

Input: $U, N_r = N_t = N, N_i = N_r/U, N_a, L, V = \{v_1, v_2, ..., v_U\}$, $\mathbf{a}_i, i \in [1, 2, ..., U]$, $\mathbf{H}$,$SNRC$

Output: $\tilde{\mathbf{a}}_L$

1: **for** $i = 1$ to U **do**
2: $\quad SNRC[i] = sum\{\sigma_n^2(diag(\mathbf{H}_{v_i}\mathbf{H}_{v_i}^H))\}$.
3: **end for**$[\sim, index] = sort(SNRC,' ascend')$.
4: **for** $i = 1$ to U **do**
5: $\quad l_i = index(i)$,$V = \{v_i, i \neq index(1)\}$.
6: **end for**
7: $\mathbf{H}_L = [\mathbf{H}_{l_U}^T, \mathbf{H}_{l_{U-1}}^T, ..., \mathbf{H}_{l_1}^T]^T$,$\mathbf{a}_L = [\mathbf{a}_{l_U}^T, \mathbf{a}_{l_{U-1}}^T, ..., \mathbf{a}_{l_1}^T]^T$.
8: get $\mathbf{B}_L$, $\mathbf{F}_L$ and $\mathbf{G}_L$ follow the procedure in Sect. 3.
9: THP process: $\tilde{\mathbf{a}}_L = \mathbf{B}_L^{-1}(\mathbf{a}_L + \mathbf{d}_L)$.

5 Simulation and Analysis

To validate our proposed algorithm in Sect. 4, we do some simulations under different antenna parameters. Considering the THP nonlinear precoding loss, $SNR = E[|a|^2/N_0][dB] + \gamma_p^2[dB]$. We use QAM for the data stream transmissions, and set $\gamma_p^2 = 1.25dB$ according to [7]. $\mathbf{H}$ is the channel matrix between the user and the base station, whose elements follow the independent and identical cyclic complex Gaussian distribution with $CN(0, 1)$.

For the first simulation, there are a base station with twelve transmit antennas and three users each of which has four antennas. Figure 3 shows the BERs of ordered and unsorted algorithm. We can see that the BER performances of user 3 and user 2 can be improved. For example, using the "best-first" sorting, user 3 at the BER 10^{-2} has 1 dB gain than unsorted scheme. When the BER is 10^{-4}, the user 2 also has a gain than unsorted scheme. The BER performance of the user 1 has some decrease, but the average BER performance of the system has been improved after the "best-first" sorting.

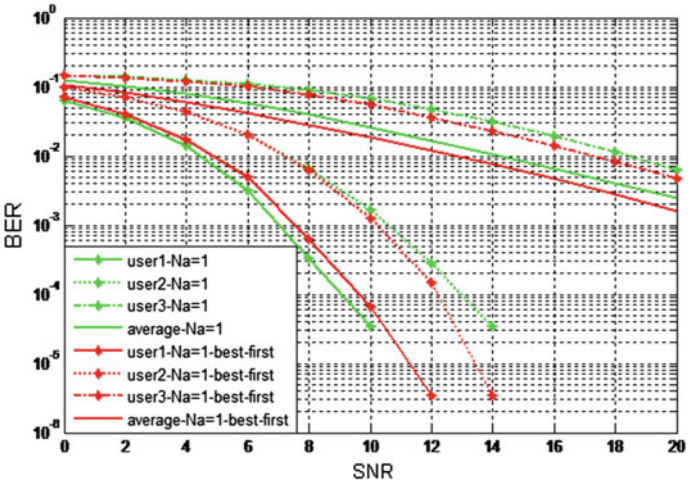


Fig. 3. 12{4,4,4} RGPSM precoding model with best-first ordering.

To verify the advantage of "best-first" sorting further, the number of base station transmitting antennas is increased to 32, the number of receiving antennas for each user is reduced to 2, and the number of users is 16. Figure 4 compares the BER performance changes of each user before and after the "best-first" sorting. Because of the large number of users, only the first two users and the last two users of the coding order are described. We can see from Fig. 4 that the system performance has useful changes after the "best-first" sorting. At the BER of 10^{-3}, the user 16 has an SNR gain of 6 dB than unsorted scheme, and the user 15 has an SNR gain of the 2 dB than unsorted scheme at the BER of 10^{-5}. The BER

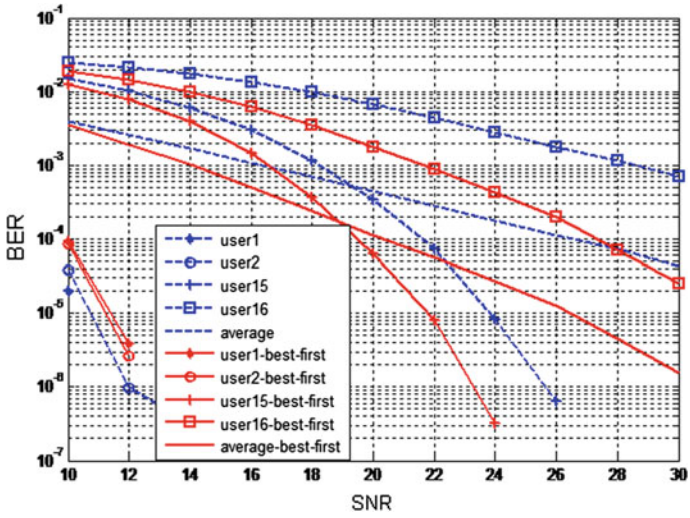


Fig. 4. 32 antennas RGPSM ($N_a = 1$) precoding model with best-first ordering.

performance of users 1, 2 has a slight decrease compared to unsorted scheme. In the case of BER at 10^{-4}, the average performance of the system gets an SNR gain near 8 dB compared to unsorted scheme. Thus, reducing the proportion of each user's data stream over total data stream can make the "best-first" sorting more obvious to the improvement of the system's overall performance.

6 Conclusions

In this paper, a novel nonlinear processing algorithm is proposed; we call it reordered generalized precoding spatial modulation (RGPSM). We studied a multiuser nonlinear precoding technique, Tomlinson–Harashima precoding (THP) precoding technique, and introduced it to GPSM system, and a single-stream system is extended to a multi-stream system to cooperate with GPSM. In addition, the "best-first" sorting method is studied to minimize BER performance difference between users. The simulation results show that the overall performance of the system is improved, and reducing the proportion of each user's data stream over total data stream can make the "best-first" sorting of the system's average performance improvement more obvious.

As for future, we can study GPSM scheme from the following aspect. In the multiuser GPSM system studied in this paper, the number of data streams received by each user is fixed. In practical applications, the base station can adjust the number of data streams that users can receive automatically according to the user's channel state and the needs of users, so as to realize a different standard GPSM and improve the overall reliability, effectiveness, and flexibility of multiuser system.

References

1. Zhang, R., Yang, L.L., Hanzo, L.: Generalised pre-coding aided spatial modulation[J]. IEEE Trans. Wirel. Commun. **12**(11), 5434–5443 (2013)
2. Zhang, R., Yang, L.L., Hanzo, L.: Error probability and capacity analysis of generalised pre-coding aided spatial modulation[J]. IEEE Trans. Wirel. Commun. **14**(1), 364–375 (2015)
3. Stavridis, A., Renzo, M.D., Haas, H.: Performance analysis of multistream receive spatial modulation in the MIMO broadcast channel[J]. IEEE Trans. Wirel. Commun. **15**(3), 1808–1820 (2016)
4. Stavridis, A., Basnayaka, D., Sinanovic, S., et al.: A virtual MIMO dual-hop architecture based on hybrid spatial modulation[J]. IEEE Trans. Commun. **62**(9), 3161–3179 (2014)
5. Li, Q., Wen, M., Li, J.: Virtual spatial modulation with diversity improvement[C]. IEEE Trans. Veh. Technol. **1**, 5 (2017)
6. Li, J., Wen, M., Cheng, X., et al.: Generalised pre-coding aided quadrature spatial modulation[J]. IEEE Trans. Veh. Technol. **99**, 1 (2016)
7. Liu, J., Krzymie, W.A.: Improved Tomlinson-Harashima precoding for the downlink of multiple antenna multi-user systems[J]. **1**(1), 466–472 (2005)
8. Spencer, Q.H., Swindlehurst, A.L., Haardt, M.: Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels[J]. IEEE Trans. Signal Process. **52**(2), 461–471 (2004)
9. Liu, J., Krzymien W.A.: A minimum mean-square error criterion based nonlinear joint transmitter-receiver processing, for the downlink of multi-user MIMO systems[C], IEEE Veh. Technol. Conf. **2006**, 1982–1987 (2006)

A Low Complexity Sphere Decoding Algorithm for Generalized Spatial Modulation

Yingze Xie, Wenbin Zhang(✉), Jinwei Huang, and Shaochuan Wu

Communication Research Center, Harbin Institute of Technology, Harbin, China
zwbgxy1973@hit.edu.cn

Abstract. There is a Sphere Decoding (SD) algorithm for Spatial Modulation (SM) (SM-Tx), which reduces the computational complexity of the detection by reducing the transmit search space in the receiver. This paper applies SM-Tx to Generalized Spatial Modulation (GSM), called as GSM-Tx, and further proposes a modified algorithm based on GSM-Tx, named as MGSM-Tx, in which a new channel real-value transformation is applied to GSM-Tx. Theoretical analysis and numerical results indicate that the Bit Error Rate (BER) and the computational complexity of MGSM-Tx depend on the system parameters, i.e., the size of the constellation and the number of active transmit antennas. Numerical results further show that MGSM-Tx is close to Maximum-Likelihood (ML) detection algorithm and GSM-Tx in BER. In addition, MGSM-Tx is always much better than GSM-Tx in reducing the computational complexity, especially at low SNR.

Keywords: MIMO · Generalized Spatial Modulation · Sphere decoding

1 Introduction

Spatial Modulation (SM) technology, as a new multiple input multiple output (MIMO) technology, has many desirable advantages, such as low complexity, low power consumption, and so on [1]. In SM, there is only one active transmit antenna at each transmission. However, SM has to request a number of transmit antennas compared with conventional MIMO. In order to overcome the limitation of the number of transmit antennas in SM, Generalized Spatial Modulation (GSM) was proposed [2]. Compared with SM, GSM can significantly improve the spectral efficiency of the system and is more applicable for MIMO [3].[1]

[1] This work was supported by the National Natural Science Foundation (NNSF) of China under Grant 61671173.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 73–81, 2019.
https://doi.org/10.1007/978-981-13-6264-4_9

Sphere Decoding (SD) algorithms for SM could reduce the computational complexity of Maximum-Likelihood (ML) detectors. A SD algorithm for SM called Transmitter-centric SD (Tx-SD) is proposed [4]. Tx-SD is designed to reduce the size of the search space at the transmitter, that is to reduce transmit search space, however, Tx-SD is not always a good solution because there is a pre-computation for the received symbol, and in some cases, Tx-SD is even more complex than ML [5]. However, GSM-Tx is better than SM-Tx in BER with the same transmit antennas [6]. Hence, it is pretty meaningful to study GSM-SD.

Motivated by mentioned considerations, this paper focuses on developing a low complexity SD detection algorithm for GSM. First, we show that the SM-Tx is not limited to SM and is applicable to GSM as well. Second, we apply a new real transformation to SM-Tx, called as GSM-Tx, and propose a modified GSM-Tx, called as MGSM-Tx. Numerical results show that MGSM-Tx has a lower computational complexity while keeping the same BER as GSM-Tx.

The remainder of this paper is organized as follows: In Sect. 2, the system model and the ML detector are summarized. In Sect. 3, GSM-Tx and MGSM detector are proposed. In Sect. 4, the computational complexity is analyzed. In Sect. 5, the simulation results are analyzed and discussed. Finally, Sect. 6 concludes the paper.

Notation: Bold lowercase and capital letters are used for column vectors and matrices, respectively. $(\cdot)^T$ denotes transposition, and the sets of real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. We use $\lfloor x \rfloor$ stand for the largest integer which is less than or equal to x. $||\cdot||$ denotes the Frobenius Norm. $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote real and imaginary parts, respectively.

2 System Model

In this paper, we consider a GSM system equipped with N_t transmit antennas, N_r receive antennas, and N_u active antennas. Without loss of generality, the information bits are mapped through an M order Quadrature Amplitude Modulation (M-QAM). Figure 1 depicts the block diagram of the GSM system. The information bits are first encoded at the transmitter. After that, $p_1 = \left\lfloor \log_2 \binom{N_t}{N_u} \right\rfloor$ bits and $p_2 = N_c \log_2 M$ bits are used to active N_u transmit antennas and implement the M-ary symbol modulation in each active antennas. Hence, the number of the transmitted bits at each transmission is $p = p_1 + p_2$. Among $\binom{N_t}{N_u}$ optional transmit antenna combinations, only $N_c = 2^{p_1}$ combinations can be used for transmission. For obtaining the close solution of formulas in the paper, we assume that only two transmit antennas are activated to transmit symbols s_m and s_l at each transmission, whose indices are m and l, respectively, where $(m, l) \in \boldsymbol{\Theta}$, and $s_m, s_l \in \{s_1, s_2, \ldots, s_M\}$. The transmit vector $\mathbf{x} \in \mathbb{C}^{N_t \times 1}$ is

$$\mathbf{x} = [\mathbf{0}_{1\times m-1}, s_m, \mathbf{0}_{1\times l-m-1}, s_l, \mathbf{0}_{1\times N_t-l}]^T \tag{1}$$

where all the entries of $\mathbf{0}_{p\times q} \in \mathbb{C}^{p\times q}$ are zero.

After the transmit vector passing through the MIMO channel $\mathbf{H}$, the received signal vector can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{2}$$

where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the complex fading channel matrix whose elements are drawn independently from the complex Gaussian distribution with zero mean and unit variance, and $\mathbf{n} \in \mathbb{C}^{N_r \times 1}$ is the additive white Gaussian noise (AWGN) with zero mean and variance σ_n^2 per dimension at the receiver.

ML-optimum detector can be written as follows:

$$\left[\hat{m}, \hat{l}, \hat{s_m}, \hat{s_l}\right] = \underset{\substack{(m,l)\in\boldsymbol{\Theta} \\ s_m, s_l \in \{s_1, s_2, \ldots, s_M\}}}{\arg\max} \{||\mathbf{y} - \mathbf{H}\mathbf{x}||_F^2\} \tag{3}$$

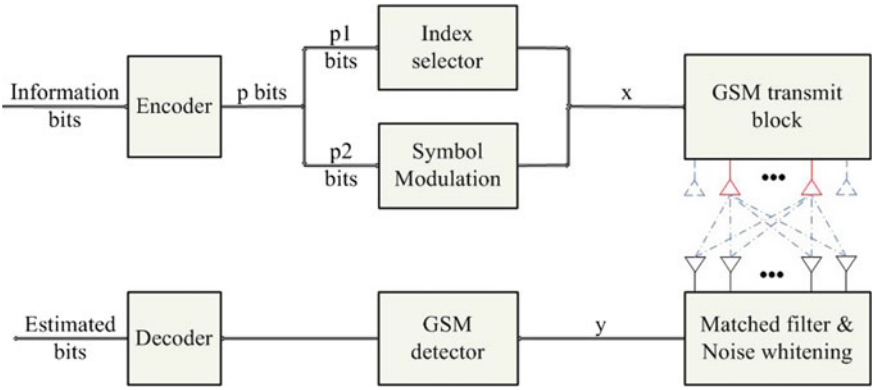


Fig. 1. Block diagram of the GSM system ($N_u = 2$).

3 GSM-Tx and MGSM-Tx Detector

3.1 GSM-Tx

Transforming (3) into expression in real field, we can obtain

$$\overline{\mathbf{y}} = \overline{\mathbf{H}}\overline{\mathbf{x}} + \overline{\mathbf{n}}, \tag{4}$$

where $\overline{\mathbf{y}} = \left[\Re\{\mathbf{y}\}^T, \Im\{\mathbf{y}\}^T\right] \in \mathbb{R}^{2N_2 \times 1}$, $\overline{\mathbf{H}} = \begin{bmatrix} Re\{\mathbf{H}\} & -\Im\{\mathbf{H}\} \\ \Im\{\mathbf{H}\} & \Re\{\mathbf{H}\} \end{bmatrix} \in \mathbb{R}^{2N_r \times 2N_t}$, $\overline{\mathbf{x}} = \left[\Re\{\mathbf{x}\}^T, \Im\{\mathbf{x}\}^T\right] \in \mathbb{R}^{2N_t \times 1}$ and $\overline{\mathbf{n}} = \left[\Re\{\mathbf{n}\}^T, \Im\{\mathbf{n}\}^T\right] \in \mathbb{R}^{2N_r \times 1}$.

Implement QR factorization for $\mathbf{H}$, and (2) can be rewritten as

$$\overline{\mathbf{y}} = \left[\overline{\mathbf{Q}}_1 \overline{\mathbf{Q}}_2\right] \begin{bmatrix} \overline{\mathbf{P}} \\ \mathbf{0}_{(2N_r - 2N_t) \times 2N_t} \end{bmatrix} \overline{\mathbf{x}} + \overline{\mathbf{n}} \tag{5}$$

where $\overline{\mathbf{P}} \in \mathbb{R}^{2N_t \times 2N_t}$ is an upper triangular matrix, $\overline{\mathbf{Q}}_1 \in \mathbb{R}^{2N_t \times 2N_r}$, $\overline{\mathbf{Q}}_2 \in \mathbb{R}^{2N_t \times (2N_r - 2N_t)}$, and define $\overline{\mathbf{z}} = \overline{\mathbf{Q}}_1^T \overline{\mathbf{y}}$ and $R_Q^2 = R^2 - ||\overline{\mathbf{Q}}_2^T \overline{\mathbf{y}}||_F^2$.

We can reduce the transmit space through the following inequalities:

$$\frac{-R_Q + \overline{\mathbf{z}}_{i+N_t}}{\overline{p}_{i+N_t, i+N_t}} \leq \overline{\mathbf{x}}_{i+N_t} \leq \frac{R_Q + \overline{\mathbf{z}}_{i+N_t}}{\overline{p}_{i+N_t, i+N_t}} \tag{6}$$

$$\frac{-R_Q^{'} + \overline{\mathbf{z}}_{j+N_t}}{\overline{p}_{j+N_t, j+N_t}} \leq \overline{\mathbf{x}}_{j+N_t} \leq \frac{R_Q^{'} + \overline{\mathbf{z}}_{j+N_t}}{\overline{p}_{j+N_t, j+N_t}} \tag{7}$$

$$\frac{-R_Q^{''} + \overline{\mathbf{z}}_i}{\overline{p}_{i,i}} \leq \overline{\mathbf{x}}_i \leq \frac{R_Q^{''} + \overline{\mathbf{z}}_i}{\overline{p}_{i,i}} \tag{8}$$

$$\frac{-R_Q^{'''} + \overline{\mathbf{z}}_j}{\overline{p}_{j,j}} \leq \overline{\mathbf{x}}_j \leq \frac{R_Q^{'''} + \overline{\mathbf{z}}_j}{\overline{p}_{j,j}} \tag{9}$$

where $i = N_t, N_t - 1, \cdots, 2$, $j = i - 1, N_t - 2, \cdots, 1$, ${R_Q^{'}}^2 = {R_Q}^2 - \sum_{k=j+N_t+1}^{i+N_t-1} (\overline{\mathbf{z}}_k)^2 - \left(\overline{\mathbf{z}}_{j+N_t} - \overline{p}_{j+N_t, i+N_t} \overline{\mathbf{x}}_{i+N_t}\right)^2$, ${R_Q^{''}}^2 = {R_Q^{'}}^2 - \sum_{k=i+1}^{j+N_t-1} (\overline{\mathbf{z}}_k)^2 - \left(\overline{\mathbf{z}}_i - \overline{p}_{i,j+N_t} \overline{\mathbf{x}}_{j+N_t}\right)^2$, and ${R_Q^{'''}}^2 = {R_Q^{''}}^2 - \sum_{k=j+1}^{i-1} (\overline{\mathbf{z}}_k)^2 - \left(\overline{\mathbf{z}}_j - \overline{p}_{j,i} \overline{\mathbf{x}}_i\right)^2$. Noting that we must calculate (6), (7), (8) and (9) in order, so the transmit search space is reduced and $\boldsymbol{\Theta}_R$ set, in which all points maintain the above inequalities, is obtained.

By (6), (7), (8), and (9), we can obtain the imaginary and real parts of two transmit symbol, respectively. Then we find a vector that lies in a certain sphere of radius R around the given vector $\widetilde{\mathbf{z}}$. However, note that (i, j) must be one of N_c combinations.

The GSM-Tx detector can be written as follows:

$$\left[\hat{m}, \hat{l}, \hat{s_m}, \hat{s_l}\right] = \underset{(m,l,s_1,s_2) \in \boldsymbol{\Theta}_R}{\arg\max} \{||\overline{\mathbf{z}} - \overline{\mathbf{P}}\overline{\mathbf{x}}||_F^2\} \tag{10}$$

However, there are two shortcomings for GSM-Tx. One is that the detection of $\overline{x}_j$ depends on $\overline{x}_{i+N_t}$, $\overline{x}_{j+N_t}$ and $\overline{x}_i$. The other is that each dimension must be searched in turn and there is no way to implement algorithm efficiently in parallel to improve the efficiency of the detection. To tackle with these two problems, we apply a new real transformation of the channel matrix $\mathbf{H}$ to GSM-Tx.

3.2 MGSM-Tx

The complex field is transformed into a real field according to new real-value transformation, so the real-valued equivalent of the channel matrix $\mathbf{H}$ and the transmit vector $\mathbf{x}$, the noise vector $\mathbf{n}$ and $\mathbf{y}$ of MGSM-Tx can be rewritten as $\widetilde{\mathbf{H}} = \begin{bmatrix} \overline{h_{(1,1)}} & \cdots & \overline{h_{(1,N_t)}} \\ \vdots & \ddots & \vdots \\ \overline{h_{(N_r,1)}} & \cdots & \overline{h_{(N_r,N_t)}} \end{bmatrix}$, $\widetilde{\mathbf{x}} = [\overline{x_1}, \cdots, \overline{x_{N_t}}\}]^T$, $\widetilde{\mathbf{n}} = [\overline{n_1}, \cdots, \overline{n_{N_t}}]^T$, and $\widetilde{\mathbf{y}} = [\overline{y_1}, \cdots, \overline{y_{N_t}}]^T$, respectively, where $h_{i,j}$ is the complex channel gain from the j-th transmit antenna to the i-th receive antenna, $i = 1, 2, \cdots, N_r$ and $j = 1, 2, \cdots, N_t$. x_j is the complex symbol transmitted by the j-th transmit antenna, n_i is the complex noise received by the i-th receive antenna and y_i is the complex symbol received by the i-th receive antenna.

Implementing QR factorization of $\widetilde{\mathbf{H}}$, we obtain

$$\widetilde{\mathbf{H}} = \left[\widetilde{\mathbf{Q}}_1 \widetilde{\mathbf{Q}}_2\right] \begin{bmatrix} \widetilde{\mathbf{P}} \\ \mathbf{0}_{(2N_r - 2N_t) \times 2N_t} \end{bmatrix} \tag{11}$$

where $\widetilde{\mathbf{P}} \in \mathbb{R}^{2N_t \times 2N_t}$ is an upper triangular matrix and $\widetilde{\mathbf{Q}}_1 \in \mathbb{R}^{2N_t \times 2N_r}$, $\widetilde{\mathbf{Q}}_2 \in \mathbb{R}^{2N_t \times (2N_r - 2N_t)}$. Define $\widetilde{\mathbf{z}} = \widetilde{\mathbf{Q}}_1^T \widetilde{\mathbf{y}}$ and $R_Q^2 = R^2 - ||\widetilde{\mathbf{Q}}_2^T \widetilde{\mathbf{y}}||_F^2$.

Being different from the first real transformation, there is a special property, which has been proven [7], for $\widetilde{\mathbf{P}}$ in the new one, as follows:

$$\widetilde{p}_{k,k+1} = 0 \tag{12}$$

where $\widetilde{p}_{k,k+1}$ is $(k, k+1)$-th entry of matrix $\widetilde{\mathbf{P}}$, and $k = 1, 3, \cdots, 2N_t - 1$.

We can find the set $\boldsymbol{\Theta}_R^*$ through the following inequalities:

$$\frac{-R_Q + \widetilde{\mathbf{z}}_{2i}}{\widetilde{p}_{2i,2i}} \leq \widetilde{\mathbf{x}}_{2i} \leq \frac{R_Q + \widetilde{\mathbf{z}}_{2i}}{\widetilde{p}_{2i,2i}} \tag{13}$$

$$\frac{-R_Q + \widetilde{\mathbf{z}}_{2i-1}}{\widetilde{p}_{2i-1,2i-1}} \leq \widetilde{\mathbf{x}}_{2i-1} \leq \frac{R_Q + \widetilde{\mathbf{z}}_{2i-1}}{\widetilde{p}_{2i-1,2i-1}} \tag{14}$$

$$\frac{-R_Q^* + \widetilde{\mathbf{z}}_{2j}}{\widetilde{p}_{2j,2j}} \leq \widetilde{\mathbf{x}}_{2j} \leq \frac{R_Q^* + \widetilde{\mathbf{z}}_{2j}}{\widetilde{p}_{2j,2j}} \tag{15}$$

$$\frac{-R_Q^* + \widetilde{\mathbf{z}}_{2j-1}}{\widetilde{p}_{2j-1,2j-1}} \leq \widetilde{\mathbf{x}}_{2j-1} \leq \frac{R_Q^* + \widetilde{\mathbf{z}}_{2j-1}}{\widetilde{p}_{2j-1,2j-1}} \tag{16}$$

where $i = N_t, N_t - 1, \cdots, 2$ and $j = N_t - 1, N_t - 2, \cdots, 1$, and we have defined ${R_Q^*}^2 = R_Q^2 - \sum_{k=2j+1}^{2i-2} (\widetilde{\mathbf{z}}_k)^2 - (\widetilde{\mathbf{z}}_{2i} - \widetilde{p}_{2i,2i})^2 - (\widetilde{\mathbf{z}}_{2i-1} - \widetilde{p}_{2i-1,2i-1})^2$. The details are shown in Algorithm 1.

Algorithm 1. Reducing the transmit search space

Require: $\widetilde{\mathbf{z}}$, $\widetilde{\mathbf{P}}$, R_Q^2,N_t, N_c, M, $\{s_1, s_2, \cdots, s_M\}$, $\boldsymbol{\Theta}$.
Ensure: $\boldsymbol{\Theta}_R^*$
for $k = 1, 2, \cdots, N_c$ **do**
  choose the k-th transmit antenna index combination from the set $\boldsymbol{\Theta}$ to get (j,i)
  if (13) and (14) are true **then**
    $R_Q^{*\,2} = R_Q^2 - \sum_{k=2j+1}^{2i-2} \left(\widetilde{\mathbf{z}}_k\right)^2 - \left(\widetilde{\mathbf{z}}_{2i} - \widetilde{p}_{2i,2i}\right)^2 - \left(\widetilde{\mathbf{z}}_{2i-1} - \widetilde{p}_{2i-1,2i-1}\right)^2$.
    if (15) and (16) are true **then**
      put $[\mathbf{0}_{2j-2}, x_1, x_2, \mathbf{0}_{2i-2j-2}, \widetilde{x}_{2i-1}, \widetilde{x}_{2i}, \mathbf{0}_{N_t-i}]^T$ in $\boldsymbol{\Theta}_R^*$.
    else
      continue
    end if
  else
    continue
  end if
end for

The MGSM-SD detector can be written as follows:

$$\left[\hat{m}, \hat{l}, \hat{s_m}, \hat{s_l}\right] = \underset{(m,l,s_1,s_2)\in\boldsymbol{\Theta}_R}{\arg\max} \{||\widetilde{\mathbf{z}} - \widetilde{\mathbf{P}}\widetilde{\mathbf{x}}||_F^2\} \tag{17}$$

where the nature of (17) is ML detection, however, it has less computational complexity than (4) due to search space reduced in the first step.

The radius can be chosen as $R^2 = 2\alpha N_r \sigma_n^2$, where α can be derived by solving the following equation:

$$\frac{\gamma\left(N_r, 2\alpha N_r\right)}{\Gamma\left(N_r\right)} = 1 - \varepsilon \tag{18}$$

where $\gamma\left(x, a\right) = \int_0^a \xi^{x-1} \exp\left(-\xi\right) d\xi$ is the lower incomplete Gamma function, $\Gamma\left(x\right) = \int_0^{+\infty} \xi^{x-1} \exp\left(-\xi\right) d\xi$ is the Gamma function, and ε is an arbitrary small value close to zero.

4 Performance Analysis

In this section, we compare computational complexity consumed by MGSM, GSM-ML, and GSM-Tx, where the complexity is measured by the average number of real multiplication and division operation. Note that the computation of all the following algorithm is based on $N_u = 2$.

The computational complexity of GSM-ML can be written as $C_{GSM-ML} = 8N_r M N_c$.

The computational complexity of GSM-Tx can be written as $C_{GSM-Tx} = C_{pre} + C_{\boldsymbol{\Theta}_R} + C_{finding}$, where C_{pre}, $C_{\boldsymbol{\Theta}_R}$, and $C_{finding}$ are the computational complexity of preprocessing, reducing the search space for $\boldsymbol{\Theta}_R$, and finding the closest point to the given point in $\boldsymbol{\Theta}_R$, respectively.

C_{pre} contains the calculation of QR factorization of the real channel matrix $\overline{\mathbf{H}}$, $\overline{\mathbf{z}} = \overline{\mathbf{Q}}_1^T \overline{\mathbf{y}}$ and $||\overline{\mathbf{Q}}_2^T \widetilde{\mathbf{y}}||_F^2$, as follows:

$$C_{pre} = C_{QR} + C_{\overline{\mathbf{z}}} + C_{||\overline{\mathbf{Q}}_2^T \overline{\mathbf{y}}||_F^2} \tag{19}$$

where $C_{QR} = \sum_{k=1}^{N} \left[2f(k) + f^2(k) + 2f^3(k) + 1\right] - N_r^3$, $f(k) = N_r + 1 - k$, $N = \min(N_r - 1, N_t)$, $C_{\overline{\mathbf{z}}} = 4N_t N_r$, and $C_{||\overline{\mathbf{Q}}_2^T \overline{\mathbf{y}}||_F^2} = (2N_r - 2N_t)(2N_r + 1)$.

$C_{\boldsymbol{\Theta}_R}$ and $C_{finding}$ can be written as:

$$\begin{aligned} C_{\boldsymbol{\Theta}_R} + C_{finding} = \sum_{(m,l)\in\boldsymbol{\Theta}} \{&(l - m + 2) N_{(l+N_t)} + (m + N_t - l + 2) N_{(m+N_t)} \\ &+ (l - m + 2) N_{(l)} + 2N_{(m)}\} + 5 \sum_{m,l\in\boldsymbol{\Theta}_R} \mathbf{card}\{\boldsymbol{\Theta}_R(m,l)\} \end{aligned} \tag{20}$$

where $N_{(l+N_t)}$, $N_{(m+N_t)}$, $N_{(l)}$, and $N_{(m)}$ are the number of the points which match the $\overline{x}_{l+N_t}$, $\overline{x}_{m+N_t}$, $\overline{x}_l$, and $\overline{x}_m$ interval, respectively, $\mathbf{card}(\cdot)$ denotes the cardinality in a set.

The computational complexity of MGSM-Tx can be written as $C_{MGSM-SD} = C_{pre} + C_{\boldsymbol{\Theta}_R^*} + C_{finding}$, where C_{pre} is the same as C_{pre} of GSM-Tx, $C_{\boldsymbol{\Theta}_R}$and $C_{finding}$ are the computational complexity of reducing the search space for $\boldsymbol{\Theta}_R^*$, and finding the closest point to the given point in $\boldsymbol{\Theta}_R^*$, respectively.

$C_{\boldsymbol{\Theta}_R^*}$ and $C_{finding}$ can be written as follows:

$$\begin{aligned} C_{\boldsymbol{\Theta}_R^*} + C_{finding} = \sum_{(m,l)\in\boldsymbol{\Theta}_R^*} \{&4N_{(2l)} + 4N_{(2l-1)} + 2(l - m) + 4N_{(2m)} \\ &+ 4N_{(2m-1)}\} + 5 \sum_{m,l\in\boldsymbol{\Theta}_R^*} \mathbf{card}\{\boldsymbol{\Theta}_R^*(m,l)\} \end{aligned} \tag{21}$$

where $N_{(2l)}$, $N_{(2l-1)}$, $N_{(2m)}$ and $N_{(2m-1)}$ are the number of the points which match the $\widetilde{x}_{2l}$, $\widetilde{x}_{2l-1}$, $\widetilde{x}_{2m}$, $\widetilde{x}_{2m-1}$ interval respectively.

5 Numerical Results

In this section, we verify the proposed detection algorithm via computer simulation and numerical results. In the computer simulations, GSM systems is considered over Rayleigh flat fading MIMO channels. Furthermore, the initial radius R is chosen as described in (25) with $\varepsilon = 10^{-6}$.

Figure 2 shows BER comparison between GSM-ML, GSM-Tx, and the proposed GSM-SD for $N_r = 8$, $N_t = 5$. From Fig. 2, it can be obviously observed that GSM-Rx, GSM-Tx, and the proposed GSM-Tx have the similar BER performance under the same modulation, and the BER of MGSM-SD with $M = 16$ is worse than that with $M = 4$. In addition, it is clear that the BER of MGSM-SD with $N_u = 2$ is similar to that with $N_u = 3$.

Figure 3 shows the relative computational complexity of the SDs with respect to the ML-optimum detector, i.e., $C_{rel} = 100 \times (C_{ML} - C_{SD})/C_{ML}$.

Figure 3 shows C_{rel} comparison between GSM-Tx and MGSM-SD for $N_r = 8$, $N_t = 5$. From Fig. 3, it is clear that MGSM-Tx is always better than GSM-Tx, especially at low SNR, and C_{rel} of all SDs with $N_u = 3$ is much larger than that with $N_u = 2$ at low SNR, however, C_{rel} of all SDs with $N_u = 3$ is similar to that with $N_u = 2$ at high SNR. In addition, it is clear that C_{rel} of all SDs with $M = 16$ is much larger than that with $M = 4$.

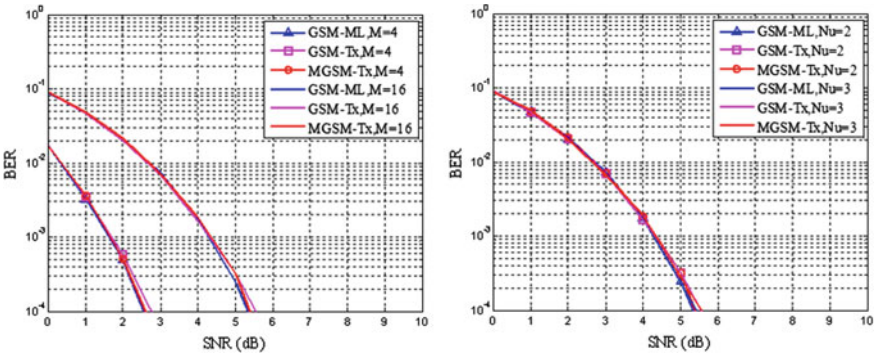


Fig. 2. BER versus the SNR, (left) $N_u = 2$, (right) $M = 16$.

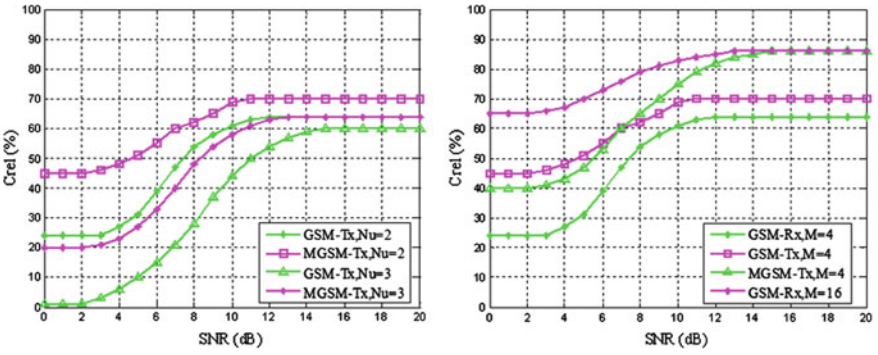


Fig. 3. Computational complexity versus the SNR, (left)$M = 4$, (right)$N_u = 2$.

6 Conclusion

In this paper, we have introduced GSM-Tx and MGSM-SD. Theoretical analysis and numerical results indicate that the computational complexity of MGSM-SD depends on the size of the constellation and the number of active transmit antennas at the transmitter. Numerical results further show that MGSM-SD is the close to GSM-ML and GSM-Tx in BER, and always much better than GSM-Tx in the computational complexity, especially at low SNR.

References

1. Renzo, M.D., Haas, H., Ghrayeb, A., et al.: Spatial modulation for generalized MIMO: challenges, opportunities, and implementation. Proc. IEEE **102**(1), 56–103 (2013)
2. Younis, A., Serafimovski, N., Mesleh, R., et al.: Generalised spatial modulation. In: Signals, Systems and Computers, pp. 1498–1502. IEEE (2011)
3. Younis, A., Mesleh, R., Renzo, M.D., et al.: Generalised spatial modulation for large-scale MIMO. In: Signal Processing Conference, pp. 346–350. IEEE (2014)
4. Younis, A., Renzo, M.D., Mesleh, R., et al.: Sphere decoding for spatial modulation. In: IEEE International Conference on Communications, pp. 1–6. IEEE (2011)
5. Younis, A., Sinanovic, S., Di Renzo, M., et al.: Generalised sphere decoding for spatial modulation. IEEE Trans. Commun. **61**(7), 2805–2815 (2013)
6. Fu, J., Hou, C., Xiang, W., et al.: Generalised spatial modulation with multiple active transmit antennas. In: GLOBECOM Workshops, pp. 839–844. IEEE (2011)
7. Azzam, L., Reduced, Ayanoglu E., Decoding, complexity sphere, for square QAM via a new lattice representation. In: Global Telecommunications Conference: GLOBECOM'07, vol. 2007, pp. 4242–4246. IEEE (2007)

A Novel Acquisition Structure for Deep Spread Spectrum System Combined with Coherent Fast Frequency Hopping

Jiapei Zhang[1(✉)], Jun Liu[2], Celun Liu[1], and Yanming Xue[1]

[1] School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China
jesprizhang@163.com
[2] Troops 66135, Beijing 100144, China

Abstract. In order to solve the problem of fast acquisition of anti-jamming measurement and control signal in deep spread spectrum system combined with coherent fast frequency hopping, a frequency sweep and gliding correlation acquisition structure of coherent DS/FFH system is proposed. Meanwhile, the gliding correlation structure is improved by proposing a Partial Matched Filtering and Fast Fourier Transform (PMF-FFT) spread spectrum PN code gliding correlation algorithm. The PMF-FFT algorithm can reduce acquisition time and FFT point to meet the real-time processing requirements and a reasonable FFT point can be selected according to Doppler frequency offset to achieve the best acquisition effect. The acquisition structure has excellent acquisition performance under strong wideband jamming and the implementation complexity is very low.

Keywords: Deep spread spectrum · Coherent fast frequency hopping · PMF-FFT · Anti-jamming

1 Introduction

The spread spectrum system has received more and more attention in recent decades, and its anti-jamming performance is considered as the most vital communication system. The coherent DS/FFH system combines the advantages of direct sequence spread spectrum system and coherent fast frequency hopping system and overcomes the shortcomings of the two spread spectrum systems when they operate independently, so it has excellent performance [1].

The acquisition methods of direct sequence spread spectrum system mainly include gliding correlation method, matched filtering method in time domain and FFT fast acquisition method in frequency domain. Reference [2] proposed a gliding correlation method, that approach is simple to implement, but acquisition time is very long when code phase difference is very large. In [3], the author proposed a matched filtering method which is suitable for occasions that Doppler frequency offset is small. But for deep spread spectrum signal, Doppler frequency offset is large (more than symbol rate before spread spectrum). The acquisition time is still very long because the acquisition

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 82–91, 2019.
https://doi.org/10.1007/978-981-13-6264-4_10

is a process of two-dimensional search in frequency offset and code phase directions. The FFT fast acquisition method proposed in [4] obtains the coarse estimate of the frequency offset while searching for the PN code phase, which greatly reduces the acquisition time, but for the deep spread spectrum signal, real-time processing requirements are difficult to meet because FFT point will become very large.

The difficulty of coherent fast frequency hopping system is frequency hopping synchronization, including the acquisition of frequency hopping sequence and the diversity combining algorithm. Frequency hopping sequence acquisition algorithms mainly include single-channel receive gliding correlation acquisition algorithm, the multichannel matched filtering acquisition algorithm, and the combination of the two algorithms [5]. The diversity combining algorithms mainly include coherent maximum likelihood combining algorithm [6], coherent linear combining algorithm [7], and coherent linear combining algorithm based on DAGC [8].

In the following sections, signal model of coherent DS/FFH system will be built first. Then, based on gliding correlation method of direct sequence spread spectrum system, single-channel receive gliding correlation acquisition algorithm and coherent linear combining algorithm of coherent fast frequency hopping system, a frequency sweep and gliding correlation acquisition structure of coherent DS/FFH system is proposed. Later, we propose a PMF-FFT spread spectrum PN code gliding correlation algorithm which can improve the structure. Lastly, the performance of the proposed acquisition structure is analyzed under strong wideband jamming.

2 Signal Model

This article studies the acquisition problem of coherent DS/FFH system. The original information sequence is multiplied by a direct sequence spread spectrum PN code to obtain direct sequence spread spectrum signal. Then the signal is modulated by PSK. The signal expression $C(t)$ after PSK modulation is (1):

$$C(t) = B_{[INT(\frac{t}{T_b})+1]} \times P_{[INT(\frac{t}{T_c})+1]} \times \cos(2\pi f_0 \times INT(\frac{t}{T_c}) \times T_c) \tag{1}$$

where B is data symbol, T_b is data symbol period, P is spread spectrum PN code, T_c is spread spectrum PN code period, f_0 is carrier frequency, and $INT()$ is rounding operation.

This signal is mixed with the hopping frequency f synthesized by the frequency synthesizer. The expression $S(t)$ of coherent DS/FFH signal is (2):

$$S(t) = C(t) \times \cos(2\pi f_{[INT(\frac{t}{T_h})+1]} t) \tag{2}$$

where T_h is frequency hopping residence time and f is the frequency point of frequency hopping.

3 Hybrid Spread Spectrum System Acquisition Structure Description

3.1 Frequency Sweep and Gliding Correlation Acquisition Structure Description

Figure 1 shows the block diagram of the frequency sweep and gliding correlation acquisition structure. Figure 2 shows the frequency sweep and gliding correlation process.

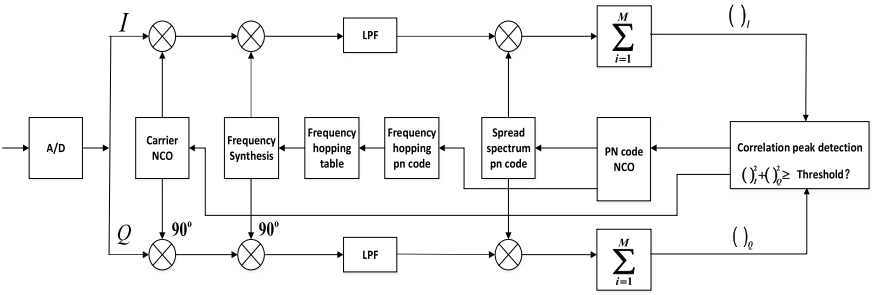


Fig. 1. Block diagram of the frequency sweep and gliding correlation acquisition structure

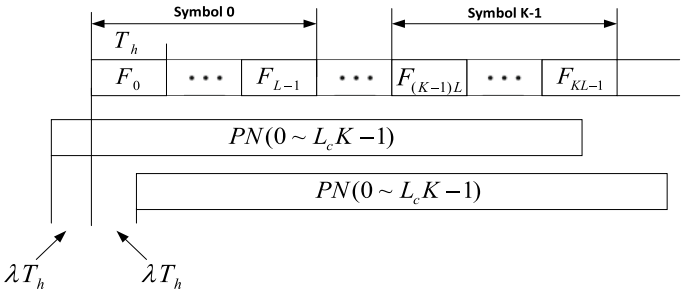


Fig. 2. Frequency sweep and gliding correlation process

The received signal enters frequency sweep process first. The frequency synthesizer generates a local frequency hopping signal according to the frequency hopping sequence. The frequency starts from the initial frequency F_0 and sweeps with frequency hopping residence time T_h. A sweep process is completed per K bits. Subsequently, it starts the next sweep process by moving the starting position of frequency sweep back $2\lambda T_h$. When the starting position of frequency moves T_b, we complete one sweep cycle.

For the baseband signal after frequency sweep process and low-pass filter(LPF), we use the spread spectrum sequence to do the gliding correlation detection. The length of the spread spectrum sequence is also K bits. For each gliding correlation process, the

start time of spread spectrum sequence is $-\lambda T_h \sim \lambda T_h$ from the initial frequency hopping position F_0, generally $\lambda \leq 0.5$.

If the correlation peak exceeds the initial threshold, the signal is considered acquired. At this time, it enters the verification acquisition process which still is a frequency sweep and gliding correlation process. In this process, the start position of the frequency hopping signal is adjusted according to the start position of PN code phase obtained in initial acquisition process, so that the transceiver signal is aligned in code phase in time, and the frequency offset is corrected according to the Doppler information obtained. When the correlation value exceeds the verification threshold, it is believed that the verification is correct and the acquisition is completed; otherwise, it is considered to be a false acquisition and the verification fails, and it is necessary to revert to the frequency sweep and gliding correlation process until the acquisition is truly completed.

3.2 PMF-FFT Spread Spectrum PN Code Gliding Correlation Algorithm Description

PMF-FFT Spread Spectrum PN Code Gliding Correlation Algorithm.
The gliding correlation PN code acquisition process not only searches for the PN code phase but also searches the carrier Doppler frequency offset. It is a two-dimensional search process. In order to reduce acquisition time, we can improve the gliding correlation structure and introduce FFT spectrum analysis. When the local PN code is consistent with the input signal PN code phase, the multiplication result is only the residual carrier frequency offset $e^{j2\pi f_d t}$, and the Doppler frequency offset value f_d can be obtained by the FFT spectrum analysis.

The acquisition structure based on PMF-FFT algorithm is shown in Fig. 3.

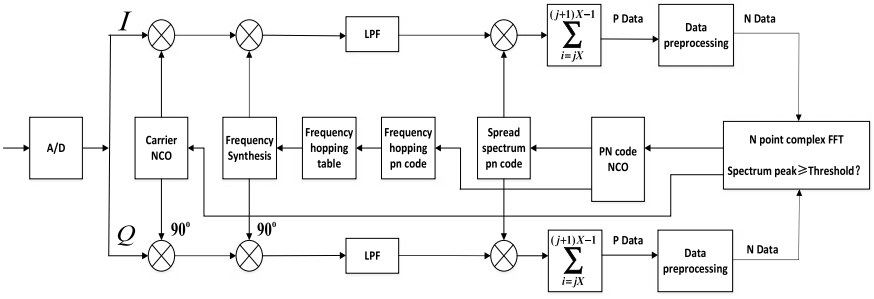


Fig. 3. Acquisition structure based on PMF-FFT algorithm

In this structure, relative slide between the local PN code and the input signal PN code still remains. After the input signal is multiplied with the local PN code, each X chip is accumulated, the input data is decelerated to $1/X$ of the original, and the original M chips becomes the current P data, $P = M/X$. This process is a partial correlation process called Partial Matched Filtering (PMF). Use this P data to do $N(N \geq P)$ point FFT operation and perform spectrum analysis. Here N is an integer power of 2, if P is not an integer power of 2, then zero-point data is added after P data, called data preprocessing.

If the peak value of spectrum exceeds the threshold value, it means that the phase of local code is consistent with the code phase of input signal, and the corresponding value of peak is the Doppler frequency offset estimate. The local carrier NCO and PN code NCO are adjusted based on the estimated value to complete the acquisition. The above structure is referred to as a PMF-FFT acquisition structure, and the main function of PMF is to perform deceleration processing on the multiplied data and reduce the point of FFT operation to meet real-time processing requirements.

Amplitude–Frequency Response of PMF-FFT Acquisition Structure.
In order to simply analyze amplitude–frequency response of PMF-FFT acquisition structure, only the situation where the local PN code and the received PN code have been synchronized, there is no noise, and only the Doppler frequency offset is left in the received signal is discussed. The amplitude–frequency response of PMF-FFT acquisition structure is (3) [9].

$$\begin{aligned} G(f_d,k) &= \frac{1}{M}\left|\frac{\sin(\pi\frac{M}{P}f_dT_c)}{\sin(\pi f_dT_c)} \times \frac{\sin(\pi M f_dT_c - \pi\frac{P}{N}k)}{\sin(\pi\frac{M}{P}f_dT_c - \pi\frac{1}{N}k)}\right| \\ &= \left|\frac{\sin(\pi\frac{M}{P}f_dT_c)}{\frac{M}{P}\sin(\pi f_dT_c)}\right| \times \left|\frac{\sin(\pi M f_dT_c - \pi\frac{P}{N}k)}{P\sin(\pi\frac{M}{P}f_dT_c - \pi\frac{1}{N}k)}\right| \\ &= G_1(f_d)G_2(f_d,k) \end{aligned} \tag{3}$$

In Eq. (3), $G_1(f_d)$ is caused by PMF, when $X = 1, P = M$, $G_1(f_d)$ has maximum value 1. $G_2(f_d,k)$ is caused by incomplete phase compensation in FFT operation, to get maximum value, we make $k = m$, where m is the nearest integer to $\frac{NM}{P}f_dT_c$. Suppose that phase compensation is optimal, we get (4):

$$G(f_d,k) = G_1(f_d)G_2(f_d,k)|_{k=m} = G_1(f_d)G_2(f_d) \tag{4}$$

Figure 4 shows amplitude–frequency response of PMF and FFT operation. Figure 5 shows amplitude–frequency response of PMF-FFT with different correlation lengths.

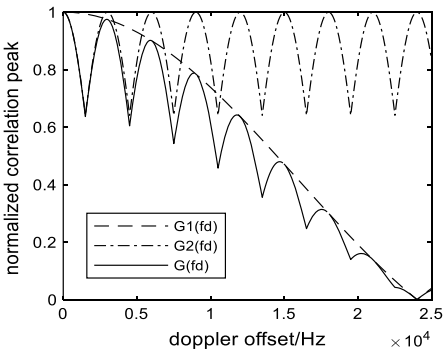


Fig. 4. Amplitude–frequency response of PMF and FFT operation

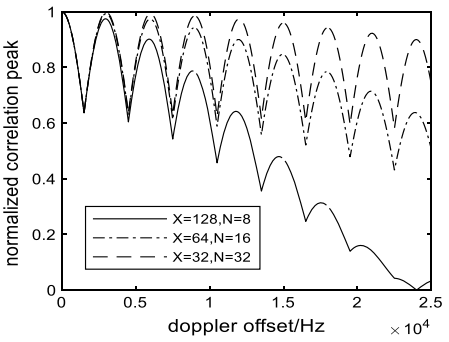


Fig. 5. Amplitude–frequency response of PMF-FFT with different correlation length

From the two figures, we can conclude that in PMF-FFT acquisition structure, larger the correlation length is, less point of FFT operation is, but the correlation peak loss is also greater and it will lead to a drop in the probability of detection. In practice, the length of correlator and the number of FFT point should be reasonably selected according to the range of Doppler frequency offset and the operation ability of device.

4 Anti-jamming Performance

For one acquisition structure, miss probability and false probability in noisy and disturbed environments are the most critical performance indicators. This section mainly analyzes miss probability and false probability of frequency sweep and gliding correlation acquisition structure under wideband jamming. In order to simplify the analysis, it is assumed that the baseband signal after de-hopping and de-spread has no code phase difference and Doppler frequency difference.

After de-hopping and de-spread, the baseband signal can be expressed as:

$$r(t) = s(t) + w(t) + j(t) \tag{5}$$

where $s(t)$ is useful signals, $w(t)$ is baseband noise, and $j(t)$ is baseband interference signals.

Each baseband signal sampling point obeys $N(\sqrt{E_b/L_c}, N_0 + N_J)$ distribution. Where E_b/L_c is the useful signal energy in a chip;N_0 is single sideband noise power spectral density; N_J is single sideband jamming power spectral density

The correlation accumulation process can be expressed as

$$R(i) = \sum_{n=i}^{i+KL_c-1} r(iT_c) = \sum_{n=i}^{i+KL_c-1} r(iMT_s) \tag{6}$$

where $M = T_c/T_s = f_s/f_c$. The correlation accumulation process is a summation process, so the correlation value still obeys the Gauss distribution.

The verdict R_i is the square of the amplitude of $R(i)$, that is:

$$R_i = |R(i)|^2 = \mathrm{Re}[R(i)]^2 + \mathrm{Im}[R(i)]^2 \tag{7}$$

4.1 Miss Probability

If the frequency hopping and spread spectrum PN code are already aligned, the relative value $R(i)$ obeys $N[K\sqrt{E_b L_c}, KL_c(N_0 + N_J)]$, so $\frac{2R_i}{KL_c(N_0+N_J)}$ obeys the non-central χ^2 distribution $nc\chi^2[2, 2KE_b/(N_0 + N_J)]$ [10], Therefore, in a sweep cycle, the miss probability P_L is

$$P_L = 1 - (1 - P_l)(1 - P_l') \tag{8}$$

$$P_l = \int_0^{\frac{2R_{th}}{KL_c(N_0+N_J)}} nc\chi^2[2, 2KE_b/(N_0 + N_J)]dx \tag{9}$$

$$P_l' = \int_0^{\frac{2R_{ck}}{KL_c(N_0+N_J)}} nc\chi^2[2, 2KE_b/(N_0 + N_J)]dx, \tag{10}$$

where P_l is miss probability in initial acquisition process, and P_l' is miss probability in verification acquisition process.

After X sweep cycles, there will be X chances to acquire useful signals, then the miss probability P_{LX} is

$$P_{LX} = P_L^X \tag{11}$$

4.2 False Probability

If the frequency hopping and spread spectrum PN code are not aligned, because of good mutual-correlation of frequency hopping and spread spectrum PN code. The correlation value $R(i)$ obeys $N[0, KL_c(N_0 + N_J)]$, so $\frac{2R_i}{KL_c(N_0+N_J)}$ obeys the $\chi^2(2)$ distribution.

Therefore, the false probability in one decision is

$$P_{f1} = \int_{\frac{2R_{th}}{KL_c(N_0+N_J)}}^{\infty} \chi^2(2)dx \tag{12}$$

Suppose that there are Y decisions in a sweep process, then the false probability in one sweep process is

$$P_f = 1 - (1 - P_{f1})^Y \tag{13}$$

When the signal is falsely acquired, it is necessary to verify the correlation result by one decision of the verification acquisition process, then the false probability is

$$P_F = P_f P_f' \tag{14}$$

$$P_f' = \int_{\frac{2R_{ck}}{KL_c(N_0+N_J)}}^{\infty} \chi^2(2)dx, \tag{15}$$

where P_f' is the probability that correlation value exceeds verification threshold R_{ck} in the verification acquisition process.

If there is Z sweep process in a sweep cycle, then the false probability P_{FZ} is

$$P_{FZ} = 1 - (1 - P_F)^Z \tag{16}$$

4.3 Simulation Results

The simulation parameters: direct sequence spread spectrum PN code spread spectrum ratio $L_c = 10000$, number of frequency hopping point $N_h = 128$, frequency diversity number of frequency hopping $L_h = 100$, frequency interval of frequency hopping R_c, number of correlation accumulation bits $K = 8$, number of sweep cycles $X = 5$, number of decisions in a sweep process $Y = 16000$, number of sweep process in a sweep cycle $Z = 200$, normalized signal-to-noise ratio $E_b/N_0 = 15$ dB, and signal-to-jamming ratio $SJR = \frac{E_b R_b}{N_J W_s} = \frac{E_b R_b}{N_J N_h L_c R_b} = \frac{E_b/N_J}{N_h L_c}$.

Figure 6 gives P_l and P_{f1} under different SJR and different initial threshold R_{th}. Figure 7 gives P_L and P_{FZ} under different SJR and different verification threshold R_{ck} when initial threshold $R_{th} = 12KL_cE_0$.

The two figures show that when the threshold is fixed, as SJR gets bigger, miss probability will decrease significantly because useful signal power will increase, and false probability is almost constant because jamming still accounts for the main energy; when SJR is fixed, as threshold gets bigger, miss probability will increase, and false probability will decrease. Thus, in practice, we need to select the appropriate threshold to make miss probability and false probability all meet the requirements.

5 Conclusion

In this paper, we have proposed a frequency sweep and gliding correlation acquisition structure of coherent DS/FFH system. The acquisition structure is very low in complexity. In order to reduce the acquisition time, we proposed a PMF-FFT spread

spectrum PN code gliding correlation algorithm. The acquisition performance of the frequency sweep and gliding correlation acquisition structure under wideband jamming is also analyzed theoretically and simulated by Matlab. The acquisition performance mainly includes miss probability and false probability. The results show that this frequency sweep and gliding correlation acquisition structure still has excellent acquisition performance under strong wideband jamming.

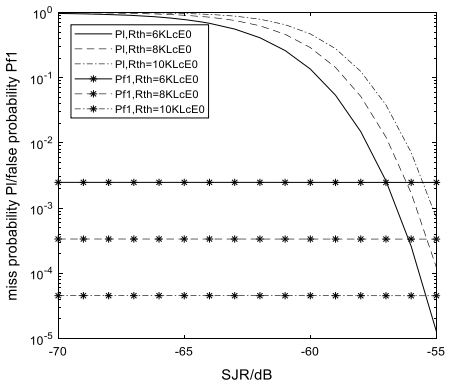


Fig. 6. P_l and P_{f1} under different *SJR* and different initial threshold R_{th}

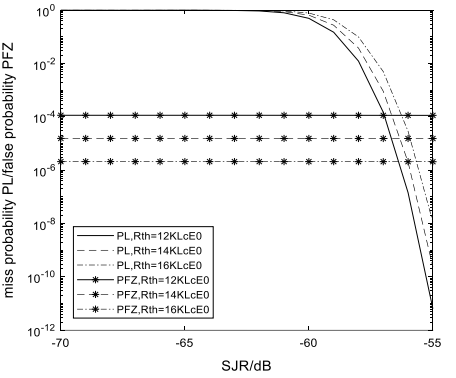


Fig. 7. P_L and P_{FZ} under different *SJR* and different verification threshold R_{ck}

Acknowledgments. The work described in this paper was supported by the National Natural Science Foundation of China under Grant 61601027.

References

1. Peterson, R.L., Ziemer, R.E.: Digital Communications and Spread Spectrum System. Macmillan, New York (1985)
2. Zhang,.S.Y., Tong, Y.J., Duan, Y.L.: Study of synchronization acquisition for PN code based on glide correlation. Inf. Commun. (2015)
3. Xiang, J.W., Wei, Z.M., Xiong, X.J.: A pseudo-code acquisition method based on digital matched filter. Radio Eng. (2011)
4. Hao, C., Li, Z., Zheng, D.: Fast acquisition method of GPS signal based on FFT cyclic correlation. Int. J. Commun. Netw. Syst. Sci. **10**(8), 246–254 (2017)
5. Putman, C.A., Rappaport, S.S., Schilling, D.L.: A comparison of schemes for coarse acquisition of frequency-hopped spread-spectrum signals. IEEE Trans. Commun. **31**(2), 183–189 (1983)
6. Li, G., Wang, Q., Bhargava, V.K., Mason, L.J.: Maximum-likelihood diversity combining in partial-band noise. IEEE Trans. Commun. **46**(12), 1569–1574 (2002)
7. Simon, M., Polydoros, A.: Coherent detection of frequency-hopped quadrature modulations in the presence of jamming–part i: QPSK and QPSK modulations. IEEE Trans. Commun. **29** (11), 1644–1660 (2003)

8. Liu, C.L., Han, L.B., Su, X., Wen, X.J.: Performance analysis of DAGC diversity combining in coherent DS/FFH system. In: IEEE International Conference on Microwave Technology & Computational Electromagnetics, pp. 376–378 (2014)
9. Spangenberg, S.M., Scott, I., Mclaughlin, S., Povey, G.J.R., Cruickshank, D.G.M., Grant, P. M.: An FFT-based approach for fast acquisition in spread spectrum communication systems. Wireless Pers. Commun. **13**(1–2), 27–55 (2000)
10. Random, Z.W., Analysis, S.: Harbin Industrial University Press, pp. 66–76. Harbin, China (1999)

Performance Analysis of Energy-Efficient Hybrid Precoding for mmWave Massive MIMO System

Mengqian Tian, Shangang Fan, Jie Yang, Jian Xiong, and Guan Gui(✉)

College of Telecommunication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
guiguan@njupt.edu.cn

Abstract. For the millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) system, hybrid precoding poses a big challenge because of its hardware cost and energy consumption and obvious performance loss in two typical architectures. One of the effective methods so-called the switch and inverter (SI)-based hybrid precoding architecture has been developed recently while lack of comprehensive performance analysis. In this paper, detailed performance analysis on energy efficiency is provided through simulation on the two-stage hybrid precoding, antenna selection (AS)-based hybrid precoding, and adaptive cross-entropy (ACE)-based hybrid precoding. It is aimed to prove the performance of ACE-based hybrid precoding is better than that of the others only with the limited ranges of values of all parameters (i.e., $N \geq 36$, $1 \leq K \leq 12$). At last, the optimal values in these parameters (i.e., $N = 36$, $K = 2$) are determined, and we prove that they can obtain the best performance.

Keywords: mmWave massive MIMO · Energy efficiency · Sum rate · Hybrid precoding · Performance analysis

1 Introduction

In recent years, as the researches on the fifth generation (5G) wireless communication systems are undertaken, the millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) technology has been considered as a promising technology to improve the achievable rate and energy efficiency [1]. The typical architectures of hybrid precoding, such as the finite-resolution phase shifters (PS)-based hybrid precoding architecture [2] and the switches (SW)-based hybrid precoding architecture [3], have some disadvantages, e.g., hardware cost and energy consumption, as well as obviously performance loss.

In order to solve aforementioned problems, a switch and inverter (SI)-based hybrid precoding architecture for massive MIMO systems has been proposed in [4]. It can reduce hardware cost and energy consumption without obvious performance loss. An adaptive cross-entropy (ACE)-based hybrid precoding scheme has been designed for

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 92–99, 2019.
https://doi.org/10.1007/978-981-13-6264-4_11

the SI-based architecture in [4]. However, the comprehensive performance analysis does not give for fully understanding the proposed scheme.

In this paper, we provide comprehensive performance analysis on the ACE-based hybrid precoding by evaluating the achievable sum rate and energy efficiency. Different from most of the existing works, the paper analyzes the performance with number of user K and antennas N first. Then, the limited range of parameters values are discussed and the best values of every parameters are determined.

2 System Model

At first, mmWave channel model is introduced briefly. Then this section illustrates the PS-based hybrid precoding architecture, SW-based hybrid precoding architecture and SI-based hybrid precoding architecture.

2.1 MmWave Channel Model

A typical mmWave massive MIMO system with hybrid precoding is considered. The base station is assumed to be consistent with N antennas and N_{RF} radio-frequency (RF) chains, which serve K active users. Each user is connected with a single antenna [5, 6]. In the paper, it is assumed that $N_{RF} = K$ [2], in order to achieve the multiplexing gains. The received signal vector $\boldsymbol{H} = [y_1, y_2, \cdots y_K]^T$ for users in this system can be presented as [7, 8]

$$\mathrm{y} = \boldsymbol{HA}X + n \tag{1}$$

$$\boldsymbol{A} = \boldsymbol{F_{RF}F_{BB}}$$

$$\boldsymbol{H} = [h_1, h_2, \cdots h_K]^H, h_k \in \boldsymbol{H}, k = 1, 2, \cdots, K$$

$$X = [x_1, x_2, \cdots x_K]^H, x_k \in X, k = 1, 2, \cdots, K$$

$$n = [n_1, n_2, \cdots n_K]^H, \ n_k \in n, k = 1, 2, \cdots, K$$

where $\boldsymbol{H}$ is the size $N \times K$ channel matrix and h_k is a $N \times 1$ channel vector between BS and the kth user. Besides, $\boldsymbol{F_{RF}} \in \mathbb{C}^{N \times N_{RF}}$ is the analog beamformer realized by analog circuit and $\boldsymbol{F_{BB}} \in \mathbb{C}^{N_{RF} \times K}$ satisfying the total transmit power constraint as $P = \|\boldsymbol{F_{RF}F_{BB}}\|_F^2$ is the baseband digital precoder, where P presents total power consumed to transmit signals. Additionally, X presents the transmitted signal vector between the BS and K active users, as is assumed that $\mathbb{E}(XX^H) = I_K$. At last, n is an additive white Gaussian noise (AWGN) vector, and it is denoted as independent and identical distributions (i.i.d) $\mathcal{CN}(0, \sigma^2)$, where σ^2 presents the noise power. n_k is the noise received by kth user.

Considering the characteristic of mmWaves in wireless communications, a geometric-based channel model is used to present the channel as [9]

$$h_k = \sqrt{\frac{N}{L_K}} \sum_{l=1}^{L_K} \alpha_k^{(l)} \mathrm{a}\left(\varphi_k^{(l)}, \theta_k^{(l)}\right), 1 \le l \le L_K \tag{2}$$

where L_K presents the number of path for the kth user. $\alpha_k^{(l)}$ denotes the complex gain for kth user, while $\varphi_k^{(l)}$ is azimuth angle of departure (AoD) and $\theta_k^{(l)}$ is elevation angle of departure of the path l for kth user [10]. $\mathrm{a}(\varphi, \theta)$ of size $N \times 1$ presents the response vector of BS antenna array. For the uniform planar array (UPA) with N elements, the array response vector with azimuth angle φ and elevation angle θ is expressed as [11]

$$\begin{aligned} \mathrm{a}(\varphi, \theta) = \frac{1}{\sqrt{N}} [1, \cdots, e^{j\frac{2\pi}{\lambda} d(m \sin(\varphi) \sin(\theta) + n \cos(\theta))}, \cdots, \\ e^{j\frac{2\pi}{\lambda} d((N_1 - 1) \sin(\varphi) \sin(\theta) + (N_2 - 1) \cos(\theta))}]^T, \\ 1 \le m < N_1, \ 1 \le n < N_2 \end{aligned} \tag{3}$$

where $N_1 \times N_2 = N$, for the uniform planar array (UPA) with N_1 and N_2 elements on horizon and vertical, respectively. Besides, λ denotes the signal wavelength, and d is the antenna spacing, where we consider $d = \lambda/2$ [12] at mmWave frequencies.

2.2 Architectures of Hybrid Precoding

2.2.1 Two Typical Architectures of Hybrid Precoding

Two typical architectures of hybrid precoding are shown in Fig. 1. In detail, Fig. 1(a) illustrates the PS-based architecture with finite-resolution phase shifters and Fig. 1(b) illustrates the SW-based architecture with switches.

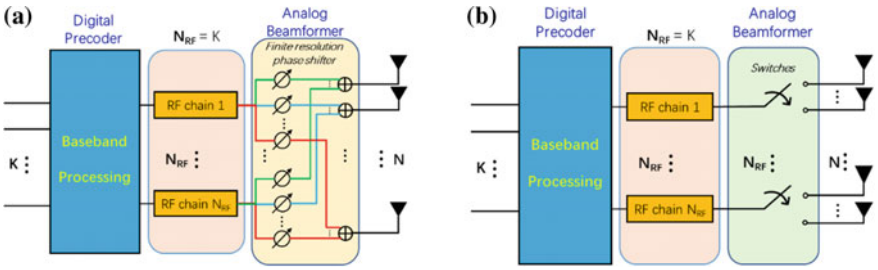


Fig. 1. Hybrid precoding: (a) typical PS-based architecture; (b) typical SW-based architecture.

As shown in Fig. 1(a), the PS-based architecture is a fully connected architecture, where each RF chain is connected to all BS antennas with a complicated phase shifter network. The PS-based architecture could acquire the near-optimal performance because of its high design freedom [2]. However, for consistent of a large number

(NN_{RF}) of finite phase shifters, it suffers from high energy consumption [12], which is present as [4]

$$P_{ps} = P + N_{RF}P_{RF} + NN_{RF}P_{PS} + P_{BB} \tag{4}$$

where P_{RF} denotes energy consumption of RF chains, as P_{PS} and P_{BB} present energy consumption of finite-resolution phase shifters and baseband, respectively.

As shown in Fig. 1(b), the SW-based architecture, where each RF chain is connected to one BS antenna, employs a small number (N_{RF}) of energy-efficient switches. Because the energy consumption of switches is much lower than finite-resolution phase shifters, it can solve the problem above. The energy consumption is expressed as follows:

$$P_{sw} = P + N_{RF}P_{RF} + N_{RF}P_{SW} + P_{BB} \tag{5}$$

where P_{SW} presents energy consumption of switches. However, only N_{RF} other than N antennas serve simultaneously so that the SW-based architecture cannot achieve the full array gains. So this architecture results in large performance loss.

2.2.2 SI-Based Architectures of Hybrid Precoding

From Fig. 2, the SI-based architecture is a sub-connected architecture, where each RF chain is connected to a subset of BS antennas with one inverter and M ($M = N/N_{RF}$) switches [13, 14]. Instead of N, the size of subset of BS antennas is M. Hence, the energy consumption is expressed as

$$P_{si} = P + N_{RF}P_{RF} + N_{RF}P_{IN} + NP_{SW} + P_{BB} \tag{6}$$

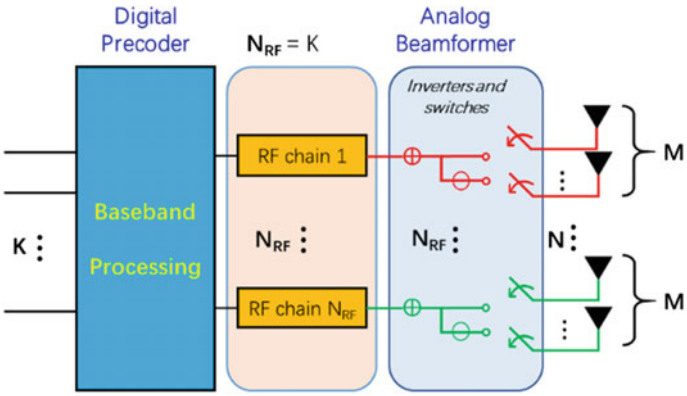


Fig. 2. Hybrid precoding: SI-based architecture.

where P_{IN} and P_{SW} denote the energy consumption of inverter and switch, respectively. Moreover, the energy consumption of inverter that can be realized by digital chin is similar to that of switch ($P_{IN} = P_{SW}$) [15].

The SI-based architecture can fully achieve the array gains of mmWave massive systems, for all antennas being used. Consequently, the energy consumption of the SI-based architecture is much lower than that of the PS-based architecture.

In this paper, the two-stage hybrid precoding scheme is designed for PS-based architecture with 4-bit finite-resolution phase shifters [2], and the AS-based hybrid precoding scheme is designed for SW-based architecture [3], and the ACE-based hybrid precoding scheme is designed for SI-based architecture [4]. Furthermore, the classical ZF digital precoder is adopted to be the digital precoder in ACE-based hybrid precoding. For each RF chain being connected with one antenna directly in the fully digital ZF precoding, the energy consumption of this scheme can be presented by

$$P_{zf} = P + NP_{RF} + P_{BB} \tag{7}$$

Table 1. Simulation parameters for all simulations

Parameters	Values
Wavelength λ	1
Antenna spacing d	$\lambda/2$
Number of path L_K	3
Complex gain $\alpha_k^{(l)}$	$\mathcal{CN}(0,1), 1 \le l \le L_K$
Azimuth angle of departure (AoD) of the path l $\varphi_k^{(l)}$	$\mathcal{U}(-\pi,\pi), 1 \le l \le L_K$
Elevation angle of departure of the path l $\theta_k^{(l)}$	$\mathcal{U}(-\pi,\pi), 1 \le l \le L_K$
Signal-to-noise ratio SNR	10 db
Total power P	0.03 W
Energy consumption of RF chains P_{RF}	0.3 W
Energy consumption of finite-resolution baseband P_{BB}	0.2 W
Energy consumption of finite-resolution phase shifters P_{PS}	0.04 W
Energy consumption of switches P_{SW}	0.005 W
Energy consumption of inverter P_{IN}	0.005 W
Number of candidates S	200
Number of elite S_{elite}	40

3 Simulation Results

In this paper, several ways of performance analysis that different from [4] will be discussed. Furthermore, the simulation results are on the performance of achievable energy-efficiency. Additionally, they are provided to analyze the performance of the ACE-based hybrid precoding by comparing to the two-stage hybrid precoding, the AS-based hybrid precoding and the fully digital ZF precoding. The computing method of sum rate is presented as follows:

$$R = \sum_{k=1}^{K} \log_2 \left(1 + \frac{\left| h_k^H \boldsymbol{F_{RF}} f_k^{BB} \right|^2}{\sum_{k'=k}^{K} \left| h_k^H \boldsymbol{F_{RF}} f_{k'}^{BB} \right|^2 + \sigma^2} \right) \tag{8}$$

where f_k^{BB} presents the kth column of $\boldsymbol{F_{BB}}$.

According to [13, 15], the energy efficiency η is defined as the ratio between the sum rate R and the energy consumption P_X of X-based architecture.

$$\eta = {}^{R}/_{P_X} \tag{9}$$

For kth user, the channel matrix is generated according to the channel model, and the simulation parameters being described in Table 1.

3.1 Performance with Number of User K

The simulation results are shown as follows.

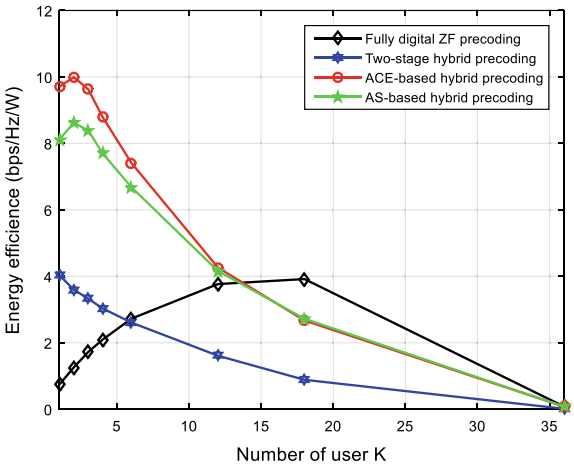


Fig. 3. Energy efficiency comparison with K.

Figure 3 shows the energy efficiency comparison in mmWave massive system, with number of elite $S_{elite} = 40$ and number of antennas $N = 36$, in addition that $N_{\mathrm{RF}} = K$ changes from 1 to 36. As shown in the figure, the energy efficiency of the ACE-based hybrid precoding is the highest, especially when K is not very large (e.g., $K \leq 12$). As K reaches 12 and more, it almost makes no difference between the energy efficiency of the ACE-based hybrid precoding and the AS-based hybrid precoding. Furthermore, when K is more than 12, the energy efficiency of the ACE-based hybrid precoding is even lower than that of fully digital ZF precoding. It indicates that this

scheme is effective when K is in a limited range, which is from 1 to 12. Last but not least, from Fig. 3(b), $K = 2$ is the best. Because it is the value of K that the ACE-based hybrid precoding can achieve best performance.

3.2 Performance with Number of Antennas N

The simulations on parameter N are done for performance analysis and the results are as the following figure.

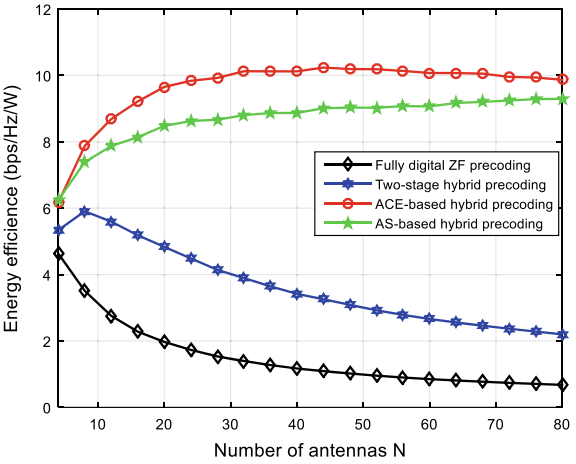


Fig. 4. Energy efficiency comparison with N.

Parameters in this simulation are similar to the simulation on the number of user, except the number of antennas N changes from 1 to 80 and the $K = 2$.

Figure 4 shows the energy efficiency comparison among four hybrid precoding schemes. Whatever, from Fig. 4, the ACE-based hybrid precoding with SI-based architecture can achieve higher energy efficiency than others. Additionally, Fig. 4 shows that the ACE-based hybrid precoding can converge with a smaller number of antennas, e.g., $N = 36$. This indicates that the number of antennas does not need to be very large, $N = 36$ is enough. These observations verify the rationality of the parameters used in Fig. 3.

4 Conclusions

In this paper, the performance of ACE-based hybrid precoding with the SI-based architecture for mmwave massive MIMO systems is analyzed. The schemes of two-stage hybrid precoding with the PS-based architecture and AS-based hybrid precoding with the SW-based architecture are incorporated in our analysis.

The performance analysis indicates that the ACE-based hybrid precoding can achieve a better performance than the other schemes, only with various parameters within a certain range. For example, when K is between 1 and 12, the ACE-based scheme can achieve higher energy efficiency than others and $K = 2$ is the best. Furthermore, when N is small, increasing N will bring about improvement in the energy efficiency, which can converge with $N = 36$. In this condition, the ACE-based hybrid precoding can achieve a considerable performance.

References

1. Zhang, J., Dai, L., He, Z., Jin, S., Li, X.: Performance analysis of mixed-ADC massive MIMO systems over rician fading channels. IEEE J. Sel. Areas Commun. **35**(6), 1327–1338 (2017)
2. Alkhateeb, A., Leus, G., Heath, R.W.: Limited feedback hybrid precoding for multi-user millimeter wave systems. IEEE Trans. Wirel. Commun. **14**(11), 6481–6494 (2015)
3. Mendez-Rial, R., Rusu, C., Alkhateeb, A., Gonzalez-Prelcic, N., Heath, R.W.: Channel estimation and hybrid combining for mmWave: Phase shifters or switches? In: Information Theory and Applications Workshop, pp. 90–97 (2015)
4. Gao, X., Dai, L., Sun, Y., Han, S., Chih-Lin, I.: Machine learning inspired energy-efficient hybrid precoding for mmWave massive MIMO systems. In: IEEE International Conference on Communications, pp. 1–6 (2017)
5. Zeng, Y., Zhang, R.: Millimeter wave MIMO with lens antenna array: a new path division multiplexing paradigm. IEEE Trans. Commun. **64**(4), 1557–1571 (2016)
6. Gao, X., Dai, L., Chen, Z., Wang, Z., Zhang, Z.: Near-optimal beam selection for beamspace MmWave massive MIMO systems. IEEE Commun. Lett. **20**(5), 1054–1057 (2016)
7. Torkildson, E., Sheldon, C., Madhow, U., Rodwell, M.: Millimeter-wave spatial multiplexing in an indoor environment. In: IEEE Globecom Workshops, Gc Workshops, pp. 1–6 (2009)
8. Xia, P., Yong, S.K., Oh, J., Ngo, C.: A practical SDMA protocol for 60 GHz millimeter wave communications. In: Asilomar Conference on Signals, Systems and Computers, pp. 2019–2023 (2008)
9. Heath, R.W., González-Prelcic, N., Rangan, S., Roh, W., Sayeed, A.: An overview of signal processing techniques for millimeter wave MIMO systems. IEEE J. Sel. Top. Signal Process. **10**(3), 436–453 (2016)
10. Sayeed, A., Brady, J.: Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies. In: IEEE Global Telecommunications Conference, pp. 3679–3684 (2013)
11. El Ayach, O., Rajagopal, S., Abu-Surra, S., Pi, Z., Heath, R.W.: Spatially sparse precoding in millimeter wave MIMO systems. IEEE Trans. Wirel. Commun. **13**(3), 1499–1513 (2014)
12. Han, S., Xu, C.I.Z., Rowell, C.: Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G. IEEE Commun. Mag. **53**(1), 186–194 (2015)
13. Gao, X., Dai, L., Han, S., Chih-Lin, I., Heath, R.W.: Energy-efficient hybrid analog and digital precoding for MmWave MIMO systems with large antenna arrays. IEEE J. Sel. Areas Commun. **34**(4), 998–1009 (2016)
14. Hur, S., Kim, T., Love, D.J., Krogmeier, J.V., Thomas, T.A., Ghosh, A.: Millimeter wave beamforming for wireless backhaul and access in small cell networks. IEEE Trans. Commun. **61**(10), 4391–4403 (2013)
15. Mendez-Rial, R., Rusu, C., Gonzalez-Prelcic, N., Alkhateeb, A., Heath, R.W.: Hybrid MIMO architectures for millimeter wave communications: phase shifters or switches? IEEE Access **4**, 247–267 (2016)

A New Multi-stage Joint Blind Equalization Carrier Synchronization for 256-QAM with Large Frequency Offsets

Chi Yonggang(✉), Qin Fengqin, Chen Lu, and Guan Yingyying

Harbin Institute of Technology, 150001 Harbin, China
chiyg@hit.edu.cn

Abstract. To solve the problem of equalization and carrier recovery in the reception and demodulation of 256-QAM with large frequency offset, this paper has proposed a multi-stage joint blind equalization carrier algorithm. This scheme first coarsely captures and compensates the signal frequency offsets, and then places the equalizer in the carrier synchronization loop to further eliminate frequency offset and inter-symbol interference. The proposed scheme greatly improves the carrier frequency offset tracking ability while eliminating the influence of carrier frequency offset on equalizer performance. Simulation results show that the proposed scheme can achieve equalization and carrier synchronization of 256-QAM with the maximum 12% normalized frequency offset

Keywords: 256-QAM · Blind equalization · Carrier recovery

1 Introduction

Because of its high spectral efficiency, quadrature amplitude modulation (QAM) has been widely used in the field of communications. 256-QAM has been adopted in DVB-C, DSL, SDH, and ITU-TJ.83 (North American cable digital TV standards). Equalization and carrier synchronization are two important steps in a high-speed QAM demodulator, which have a decisive influence on the demodulation performance. They are connected tightly.

However, most of the literature for blind equalization are based on the premise of perfect carrier synchronization, and the discussion on carrier synchronization is without considering the impact of multipath channels on reception [1, 2]. In fact, when we compensate to the channel distortion, the presence of the carrier frequency offset will make the equalizer's steady-state jitter larger and convergence to the coefficients slower or even failed. Similarly, the presence of ISI will make the carrier synchronization loop have larger steady-state error, or even cannot locked. Only the blind channel equalization has removed serious inter-symbol interference, carrier synchronization to work properly. Therefore, it is necessary to consider channel equalization and carrier synchronization at the same time [3]. The literature [4] pointed out that the MMA algorithm itself has a certain phase tracking capability which is equivalent to a first-order phase-locked loop (PLL). As we all know, the tracking bandwidth of a first-order PLL

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 100–107, 2019.
https://doi.org/10.1007/978-981-13-6264-4_12

is narrow, and there is a steady-state phase difference at the same time. The joint carrier recovery blind equalization algorithm proposed by the literature [5, 6] only uses the phase tracking ability of the MMA algorithm itself, which is not significantly different from the single MMA algorithm. In order to eliminate the phase difference and ensure the performance of the equalizer, the literature [7] proposed a new blind equalization algorithm, but the phase tracking capability is still weak. The literature [8] reduces the complexity of the algorithm. The literature [9] proposes to place the MMA equalizer in the second-order digital phase-locked loop structure, and uses the result feedback of the phase-locked loop to eliminate the carrier frequency offset of the pre-equalization data and achieves the ideal blind equalization performance without frequency offset.

The common problem of the joint structure of blind equalization and carrier synchronization is that the frequency offset acquisition range is small (approximately 1% of the symbol rate). Sometimes, the receiver needs to work at a larger carrier frequency offset. In the following, this paper proposes a new joint structure that achieves blind equalization and carrier recovery with large frequency offset of 256-QAM.

2 The Proposed Scheme

In order to blind equalize and carrier synchronize properly for 256-QAM at high rate with high-frequency offset, a joint multilevel blind equalization and carrier recovery scheme is presented in this paper, as shown in Fig. 1. First, coarse synchronization algorithm is used to coarsely capture and compensate the carrier frequency offset, and the residual carrier frequency offset is reduced to less than 1% of the symbol rate. Then the blind equalization and dual-mode carrier recovery combined structure are used to finally accurately estimate the carrier frequency and restore the correct constellation.

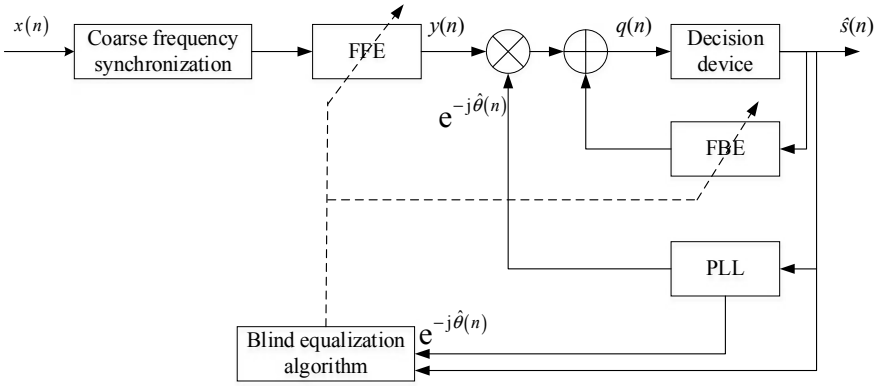


Fig. 1. The multi-stage joint blind equalization carrier algorithm

The proposed scheme can be divided into two stages.

Stage I: This stage is used to the equalization blind startup and carrier capture. At this time, the feed-forward equalizer works because of its fast convergence speed, and

the carrier synchronization loop needs an acquisition algorithm with a wide equivalent noise bandwidth. Such that the joint scheme can eliminate the inter-symbol interference and capture the frequency offset quickly.

Stage II: This stage is for the equilibrium stability as well as the carrier tracking. When the output error is less than the set threshold, the feedback equalizer begins to work, and PLL switches to a narrower equivalent noise bandwidth and a higher precision tracking algorithm. This phase further eliminates ISI, compensates the frequency offset and phase deviation, and improves the steady-state performance of the system.

2.1 Blind Equalization Algorithm

In order to take account of convergence speed and accuracy, a dual-mode blind equalization algorithm combining constant modulus algorithm (CMA) and decision-directed algorithm (DD) are adopted in this paper. In the joint processing of equalization and carrier synchronization discussed here, the selection of the blind equalization algorithm should ensure that the convergence of the equalization algorithm is not affected by the carrier frequency offset. That is, before the carrier recovery is completed, the equalizer can open the eye of the received signal.

The CMA algorithm, proposed by Godard [10], is the most widely used algorithm for blind equalization. The CMA cost function is given by

$$\begin{aligned} J_{\mathrm{CMA}}(n) &= E\left[\left(|y(n)|^2 - R_2\right)^2\right] \\ R_2 &= \frac{E\{|s(n)|^4\}}{E\{|s(n)^2|\}}, \end{aligned} \tag{1}$$

where $y(n)$ is the equalizer output, $y(n)$ is the transmitted data sequence, and $E[\bullet]$ denotes expectation. The error function of CMA is defined as

$$e_{CMA}(n) = y(n)\left(|y(n)|^2 - R_2\right) \tag{2}$$

The tap weights vector of an adaptive equalizer are adjusted by using a stochastic gradient algorithm, i.e., the complex tap vector w(n + 1) is updated according to

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu e^*_{CMA}(n)\mathbf{r}(n) \tag{3}$$

where μ is a small number called step size, and $\mathbf{r}(n)$ is input signal vector.

For a convex cost function, the tap updating algorithm has converged when the gradient is zero, that is $e(n) = 0 \Rightarrow y_R^2(n) + y_I^2(n) = R_2$ which means CMA tries to minimize the dispersion of the equalizer output $y(n)$ around the circle with a radius of $\sqrt{R_2}$. Notice that this cost function is only related to the modulus of the signals rather than the carrier phase, Therefore, whether the equalizer converges is independent of the carrier phase error. So, the CMA algorithm is not disturbed by the carrier frequency offset, can effectively open the signal eye diagram, and meets the blind startup conditions noticed before.

However, the accuracy of CMA algorithm is low. When the steady-state error becomes relatively small, we will switch to the DD algorithm to reduce the residual error.

DD blind equalization algorithm is a simple and effective blind equalization algorithm. In DD algorithm, the cost function that is usually used is the mean square error (MSE) defined as

$$J_{DD}(n) = MSE = E[e_{DD}^2(n)] = E\left[|\hat{s}(n) - y(n)|^2\right] \tag{4}$$

The error function of DD is defined as

$$e_{\mathrm{DD}}(n) = \hat{s}(n) - y(n) \tag{5}$$

The weight vector of CMA is updated by

$$\begin{aligned} \mathbf{w}(n+1) &= \mathbf{w}(n) + \mu e_{\mathrm{DD}}^*(n)\mathbf{r}(n) \\ &= \mathbf{w}(n) + \mu[\hat{s}(n) - y(n)]^*\mathbf{r}(n) \end{aligned} \tag{6}$$

DD is converged with a so-called ideal reference. So, this algorithm is simple and the residual error is small. Thus, if the DD is used in a feedback equalizer after CMA convergence has been completed, the steady-state performance of the system can be effectively improved.

2.2 Carrier Synchronization Algorithm

The coarse synchronization algorithm should have a large range of frequency offset estimation, which can correct most of the frequency offsets without introducing new errors to subsequent algorithms. The literature [11] proposes a Viterbi frequency offset estimation algorithm that uses the four constellation points on the outermost corner of the signal constellation to extract frequency offset information. The modulation information is removed from the received signal using its l power, and then the influence of the constant phase error is removed through the conjugate differential delay. And then, the frequency deviation was estimated. The specific frequency offset calculation formula of the Viterbi frequency offset estimation algorithm is

$$\Delta\hat{f} = \frac{1}{2\pi MT}\arg\{\sum_{k=0}^{L-1}\left|z_k z_{k-1}^*\right|^l e^{jM\ \arg\{z_k z_{k-1}^*\}}\} \tag{7}$$

According to the theoretical derivation, the frequency offset capture range of the Viterbi frequency synchronization algorithm can be obtained. Since the value of the phase angle of the inverse trigonometric function is always within the range $(-\pi, \pi)$, the maximum frequency deviation captures the range of the Viterbi algorithm $\Delta f \leq \pm 1/8T$. The phase capture range is $-\pi/4 \leq \hat{\theta} \leq \pi/4$.

Its carrier synchronization performance is related to the value of l.

Although the Viterbi algorithm can eliminate most of the frequency offsets, the remaining frequency offsets are relatively large. Therefore, the precise synchronization algorithm is required to further eliminate the frequency and phase offset.

Considering the factors of capture bandwidth and steady-state phase jitter, the carrier synchronization loop is divided into capture state and tracking state. At the beginning, the loop works in the capture state with a larger equivalent bandwidth to make sure that the larger frequency deviation can be captured quickly. When the error reaches a certain threshold, it enters the tracking state with a smaller loop bandwidth to reduce the steady phase jitter. In the tracking state, the direct decision loop (DD-PLL) algorithm is used.

The DD-PLL uses a decision-based phase detector. First, the current input signal $r(n)$ is compensated to obtain the signal $q(n)$ by using the phase error output value of the loop output last time. Then, the phase error is calculated by sine discriminating the $q(n)$ and its corresponding decision result $d(n)$. Next, the feedback of the loop filter (LF) and the numerically controlled oscillator (NCO) is used to the phase compensation on the next input signal $r(n+1)$, and so on. Therefore, the DD algorithm has higher accuracy within its estimation range.

The phase discrimination output $p(n)$ of DD-PLL is

$$p(n) = \mathrm{Im}\left[\frac{q(n)}{d(n)}\right] \tag{8}$$

Since the phase error is calculated by the phase difference between the $q(n)$ and its $d(n)$, the DD-PLL will fail to converge when the frequency deviation or the phase deviation of the input signal $r(n)$ is too large. So the carrier capture range of DD-PLL is small. In general, increasing the loop bandwidth can speed up the convergence of the loop and increase the frequency offset estimation range. However, the residual phase jitter also increases at the same time, that is, the increase of the frequency offset capture range is at the expense of loss accuracy.

Therefore, the PFD-PLL algorithm is used in the capture state which has a larger capture range. Assume that the carrier synchronization signal is $r(n)$. If $\tau_1 < |r(n)| < \tau_2$, the output phase error $p(n)$ remains unchanged. If not, we will make polar decisions on $p(n)$ to get $d(n)$ and calculate phase error $p(n)$. That is:

$$d(n) = \frac{|r(n)|}{\sqrt{2}}\{\mathrm{sgn}\{\mathrm{Re}[r(n)]\} + j \cdot \mathrm{sgn}\{\mathrm{Im}[r(n)]\}\} \tag{9}$$

$$p(n) = \begin{cases} p(n-1), \tau_1 < |a(n)| < \tau_2 \\ \mathrm{Im}\left\{\frac{q(n)}{d(n)}\right\}, others \end{cases} \tag{10}$$

where $q(n)$ is the phase-compensated $r(n)$. Since the output phase of polarity decision can only be $\pm 45^\circ$ or $\pm 125^\circ$, in order to prevent the phase rotation between adjacent symbols caused by too large frequency offset is too fast, the threshold of the track and hold algorithm should be 45°, when the current phase error output is calculated. When the difference between $p(n)$ and the actual output of the previous phase detector

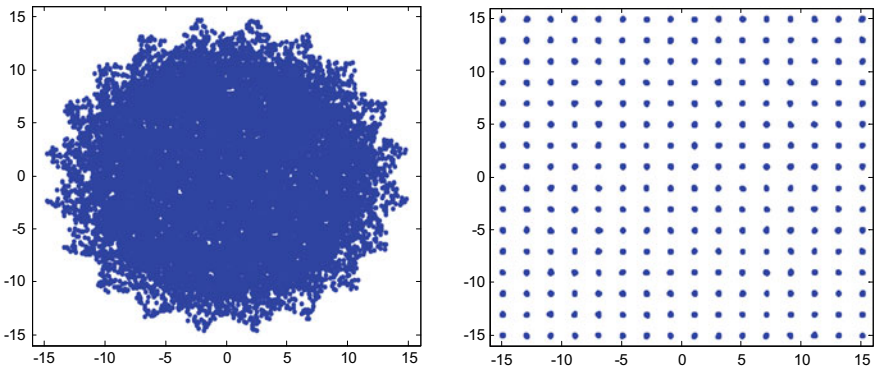
$z_{PFD}(n-1)$ is smaller than the threshold, $p(n)$ is directly output; otherwise, the phase detector output value is 45°, and its polarity is the same as the last phase detector output, i.e.: The larger the threshold τ, the smaller the steady-state phase error and the smaller the probability of sign error. Ultimately, the number of symbols required for PLL convergence, the accuracy, and the frequency offset estimation range are determined by the equivalent noise bandwidth and the threshold.

$$z_{PFD}(n) = \begin{cases} p(n), |p(n) - z_{PFD}(n-1)| < \frac{\pi}{4} \\ \frac{\pi}{4}\mathrm{sgn}[z_{PFD}(n-1)], others \end{cases} \tag{11}$$

In order to obtain faster convergence speed and larger frequency offset capture range, PFD-PLL should be used first to converge fast under a large frequency deviation. When the PFD-PLL reaches a steady state, the loop is converted to the DD-PLL mode in order to reduce residual error jitter and further improve the carrier frequency offset estimation accuracy.

3 Simulation Results

In order to analyze the performance of the proposed scheme, we present some simulation results using MATLAB. We consider 256-QAM transmissions at a symbol rate of 100Msps and the roll-off factor is 0.75. The FFE and FBE are all symbol rate spaced. The FFE is 17 taps and the FBE is 12 taps. The CMA step size is 1e-8 and the DD is 2e-5. The equivalent noise bandwidth of the PFD-PLL loop is 0.005 of the symbol rate, and the DD-PLL loop equivalent noise bandwidth takes 0.001 of the symbol rate.



(a) Input signals (b) Output signals

Fig. 2. The constellations of the proposed scheme

Figure 2 shows the convergence of the proposed scheme. The simulation parameters are 256-QAM, 40db SNR and frequency offset 5% of the symbol rate. Figure 3 shows the performance of the proposed scheme with 5%, 12% and zero normalized frequency offset. Figure 4 shows normalized frequency offset capture range of the proposed scheme and the Viterbi algorithm. It indicates that the frequency offset capture range of the proposed scheme is about 12%, which is very close to that of the Viterbi algorithm.

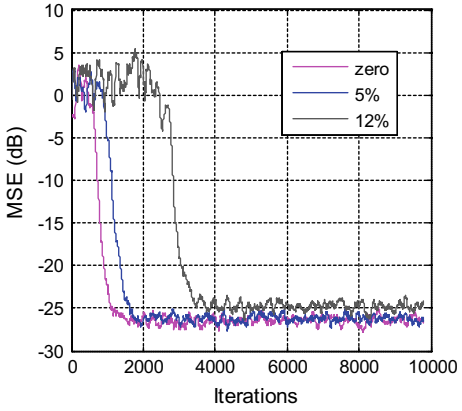


Fig. 3. The performance in different normalized frequency offsets

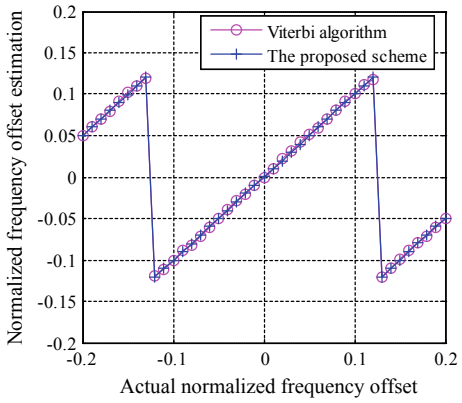


Fig. 4. The frequency offset capture range

4 Conclusion

In demodulation and reception of 256QAM signals, channel equalization and carrier synchronization are two key links. These two links are related to each other and affect each other, and they need to be jointly designed and implemented. In this paper, a multi-stage joint blind equalization carrier algorithm scheme is proposed, which realizes the blind equalization and carrier synchronization of 256-QAM signal under large

frequency offset. In the first step, the Viterbi algorithm is used to coarsely capture and compensate the signal frequency offset. The second step uses CMA algorithm and PFD-PLL to capture the frequency offset and converge the algorithm. After the system is stable, it switches to the DD algorithm and reduces the equivalent bandwidth, further reducing the steady-state error. This scheme improves the system's carrier frequency tracking ability and achieves the equalization and carrier recovery of 256-QAM with high-rate and large-frequency offset. Simulations show that this scheme increases the frequency offset range to about 12% of the symbol rate.

References

1. Park C., Lee, J., Kim, Y.: Modified reduced constellation PLL for higher order QAM [C]. In: IEEE International Symposium on Circuits and Systems, 2144–2147 (2007)
2. Xue, W., Yang, X.N., Zhang Z.Y.: A carrier recovery algorithm for high order QAM signals [C]. In: 2010 6th International Conference on Wireless Communications Networking and Mobile Computing, pp. 1–4 (2010)
3. Jablon, N.K.: Joint blind equalization, carrier recovery and timing recovery for high-order QAM signal constellations [C]. IEEE Trans. Signal Process. **40**(6), 1383–1398 (1992)
4. Yuan, J.T., Tsai, K.D.: Analysis of the multimodulus blind equalization algorithm in QAM communication systems [J]. IEEE Trans. Commun. **53**(9), 1427–1431 (2005)
5. Jablon, N.K.: Joint blind equalization, carrier recovery and timing recovery for high-order QAM signal constellations [J]. IEEE Trans. Signal Process. **40**(6), 1383–1398 (1992)
6. Zhang, Y., Zhao, J., Li, J., et al.: A fast convergent algorithm for joint blind equalization and carrier recovery[C]. In: International Conference on Signal Processing, pp. 1784–1787. IEEE (2008)
7. Abrar, S., Nandi, A.K.: Adaptive Solution for Blind Equalization and Carrier-Phase Recovery of Square-QAM [J]. IEEE Signal Process. Lett. **17**(9), 791–794 (2010)
8. Hung, K.C., Lin, D.W.: Joint carrier recovery and multimodulus blind decision-feedback equalization under high-order QAM[C]. In: Global Telecommunications Conference, GLOBECOM 2004, vol. 4, pp. 2281–2285. IEEE (2004)
9. Tian, J., Shen, B., Li, Z., et al.: Joint carrier recovery and adaptive equalization for high-order QAM [C]. In: IEEE International Symposium on Circuits and Systems, vol. 2, pp. 928–931. IEEE (2005)
10. Godard, D.N.: Self-recovering equalization and carrier tracking in two-dimensional data communication systems [J]. IEEE Trans. Commun. **28**(11), 1867–1875 (1980)
11. Aghamohammadi, A., Meyr, H., Ascheid, G.: A new method for phase synchronization and automatic gain control of linearly modulated signals on frequency-flat fading channels [J]. IEEE Trans. Commun. **39**(1), 25–29 (2001)

High-Throughput Implementation of QC-LDPC on RaPro Prototyping Platform

Bin Han[1], Fan Cao[1], Xuanxuan Gao[1], Senjie Zhang[2], and Shi Jin[1(✉)]

[1] National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
{seuhanbin,jinshi}@seu.edu.cn

[2] Key Laboratory of Universal Wireless Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract. Quasi-Cyclic Low-Density Parity-Check (QC-LDPC) codes have been adopted for the data channel in 5G New Radio (NR). This paper addresses the software-based implementation of the QC-LDPC codes described in the 5G specification. Both encoder and decoder are implemented on general purpose processor (GPP) with SIMD instructions set, and over 100 Mbps throughput is achieved on our RaPro prototyping testbed. Some scenario tests are performed and the results prove the feasibility of the proposed implementation. Time cost analysis indicates the practical use and real-time processing ability of the implemented QC-LDPC on RaPro.

Keywords: QC-LDPC codes · Encoder · Decoder · SIMD instructions set · General purpose processor · RaPro prototyping testbed

1 Introduction

Low-Density Parity-Check (LDPC) codes, proposed by Gallager in 1962 [1], have been adopted for the data channel in 5G New Radio (NR) for their near-Shannon-limit performance. It is valuable to give a flexible reference implementation of the Quasi-Cyclic Low-Density Parity-Check (QC-LDPC) codes adopted in the 5G specification. Most of the previous QC-LDPC implementations are based on ASIC and FPGA technology [2], but it is extremely time-consuming considering the development cycle. More and more works focus on the implementation of programmable QC-LDPC decoders on GPU and CPU due to the rise of parallel architectures [3]. These works demonstrate that multi-core and many-core technology can provide a flexible alternative to the implementation of LDPC. Recently, a novel 5G rapid prototyping (RaPro) testbed has been proposed in [4], which makes it easier and more flexible to implement and verify various algorithms on general purpose processor (GPP) platform. In this paper, a high-throughput implementation of QC-LDPC encoder and decoder is given based on

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 108–115, 2019.
https://doi.org/10.1007/978-981-13-6264-4_13

GPP platform with Intel SIMD instructions set, and over 100 Mbps throughput is achieved on the RaPro prototyping testbed. Some scenario tests are performed and the results show the feasibility of the proposed QC-LDPC implementation.

The subsequent contents are organized as follows. Section 2 gives a general QC-LDPC code description, and discusses some encoding and decoding algorithms of QC-LDPC. Section 3 presents the specific implementation of the QC-LDPC adopted in 5G NR. Section 4 performs some experiments on the RaPro platform and analyzes the results. Section 5 provides the concluding remarks.

2 QC-LDPC Codes

QC-LDPC codes form an important subclass of LDPC codes. A QC-LDPC code is characterized by the parity-check matrix composed of small square blocks which are the zero matrix or circulant permutation (right-shifted identity) matrices. These codes have encoding advantage over other types of LDPC codes.

2.1 Encoding of QC-LDPC

Since LDPC codes are defined by parity-check matrix $\mathbf{H}$ and the generator matrix G is usually unknown, encoding with low complexity is not straightforward. A linear encoding algorithm of QC-LDPC codes was proposed in [5], where a linear algorithm was presented to find the generator matrix of a QC-LDPC code in a systematic-circulant (SC) form, based on its parity-check matrix in circulant form. Then, the encoding of QC-LDPC codes can be accomplished with an array of shift registers in linear time complexity. This paper proposes a software implementation of this encoding algorithm based on the GPP platform with Intel SIMD instructions set.

2.2 Decoding of QC-LDPC

The most effective method for decoding LDPC codes is the belief propagation (BP) algorithm, which is a message-passing algorithm that updated iteratively on Tanner graph. A Tanner graph is a bipartite graph derived from parity-check matrix $\mathbf{H}$. The two sets of nodes in a Tanner graph are called check nodes (CNs) and variable nodes (VNs).

LP_j is the Log Likelihood Ratio (LLR) received by VN_j. Lq_{ji} is the message sent from VN_j to CN_i, and Lr_{ij} is the message sent from CN_i to VN_j. Equations (1) and (2) detail the calculation of Lq_{ji} and Lr_{ij}, where $M(j)\backslash i$ is the set of CN connected to VN_j except the CN_i, and $N(i)\backslash j$ is the set of VN connected to CN_i except the VN_j.

$$Lq_{ji} = LP_j + \sum_{i' \in M(j)\backslash i} Lr_{i'j} \ . \tag{1}$$

$$Lr_{ij} = 2\tanh^{-1}\left(\prod_{j' \in N(i)\backslash j} \tanh\left(\frac{Lq_{j'i}}{2}\right)\right) \ . \tag{2}$$

Due to the high cost of the calculation of $tanh$ in Equation (2), a simplified BP algorithm called Min-Sum (MS) algorithm was proposed in [7]. In MS algorithm, the Lr_{ij} update equation is given by

$$Lr_{ij} = \left[\prod_{j' \in N(i) \setminus j} sign(Lq_{j'i}) \right] \cdot \left[\min_{j' \in N(i) \setminus j} |Lq_{j'i}| \right] . \tag{3}$$

3 Vectorized Implementation on RaPro

In this section, an encoder and a decoder of QC-LDPC are implemented on GPP platform with Intel SIMD instructions set. According to the encoding algorithm proposed in [5], the key of the encoder is to implement a cycle shift register. An efficient decoder based on MS algorithm is given in this section, where data structure and parallelism scheme are discussed in detail.

3.1 Vectorized Encoder

This paper adopts the approved QC-LDPC scheme in 3GPP TS 38.212 [6] for 5G NR. The specification gives two base graphs for QC-LDPC parity-check matrix **H** and sets of QC-LDPC lifting size Z. Proper parameters should be selected according to the information block size and coding rate. Some critical parameters in the implementation are given in Table 1.

Table 1. The critical parameters of the QC-LDPC encoder

Parameter	Variable	Value
Information block size	K	2560
Coding rate	R	2/3
Lifting size	Z	256
Parity-check matrix size	**H**	(2560, 4352)

The encoding algorithm proposed in [5] is shown in Fig. 1, and the key of the encoder is to implement a cycle shift register. In this paper, an AVX_$m256i$ variable G_{qc} is used as a cycle shift register, and P is a variable used to store the parity-check bits.

Let **a** be the information sequence to be encoded. Every time one bit of **a** is moved into the encoder, and an AND operation is performed with G_{qc}. Then, an XOR operation is performed with P and the result is stored in P. Finally, a right cyclic shift is performed on G_{qc}.

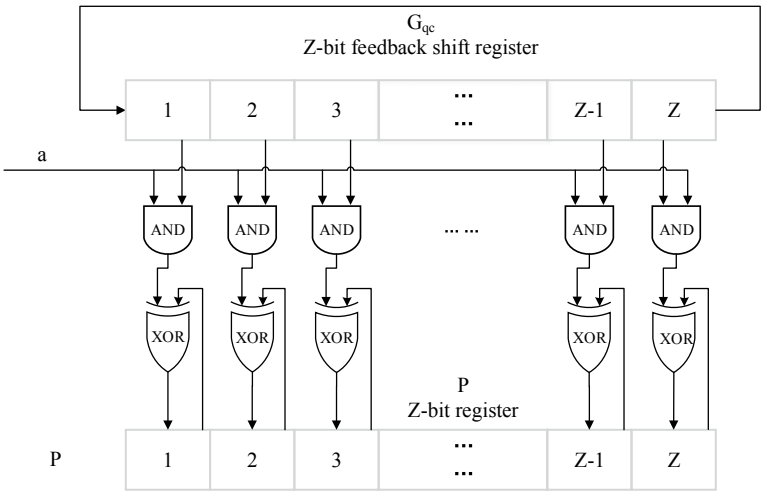


Fig. 1. The encoder of QC-LDPC based on cycle shift register

3.2 Vectorized Decoder

A direct mapping of the Tanner Graph to Lq matrix and Lr matrix is inefficient due to the scale and sparsity of **H**. An efficient decoder should take the sparsity of parity-check matrix **H** into account. Only clever usage of the instructions set functionality and exploitation of the different regions within the memory hierarchy is guaranteed to optimize the LDPC decoders for performance and efficiency of computation [7].

Data Structure. The decoder of LDPC keeps passing information iteratively between VNs and CNs, so it is vital to adopt appropriate data structure to store Lq matrix and Lr matrix for the performance of decoding.

The **H** matrix is the adjacency matrix of the Tanner graph, whose edges represent the bidirectional flow of messages exchanged between VNs and CNs. A considerable amount of memory may be required to store such structure if mapping the Tanner Graph to Lq matrix and Lr matrix directly. Therefore, a compact representation of the Tanner Graph is adopted in this paper, which is also the mainstream scheme in some related works based on GPU [8,9].

Figure 2 shows the compact method used to map the **H** matrix to a one-dimensional vector. The Lq vector is generated by scanning the **H** matrix in a row-major order and by sequentially mapping only the VN edges associated with non-null elements in **H** used by a single CN equation (in the same row). A look-up table (LUT) named V2C is used to record the index of CN, so that the Lq message of a certain CN can be read continuously and the Lr index can be obtained by the V2C table. Similarly, the Lr vector can be generated by a column-major order scan and a V2C LUT is used to store the relevant Lq index. In this way, the efficiency of data access can be improved because of the lower memory footprints of this storage scheme.

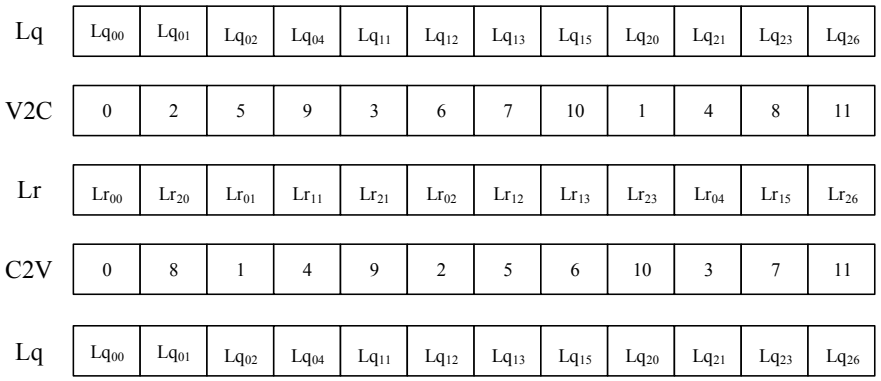


Fig. 2. The compact representation of Lq and Lr

Parallelism Scheme. To achieve high-throughput performance under programmable targets, a software LDPC decoder has to exploit computation parallelism in order to take advantage of multi-core architectures.

On account of the irregular interconnection of the Tanner graph between CNs and VNs, it is not intuitive to propose a parallelism scheme in a single codeword. To take advantage of SIMD properties, the most efficient approach consists in processing different frames in parallel [10]. Intel AVX instructions can process 8 floating-point or 16 fixed-point computations during one clock cycle, which makes it highly suitable for multiple codewords decoding.

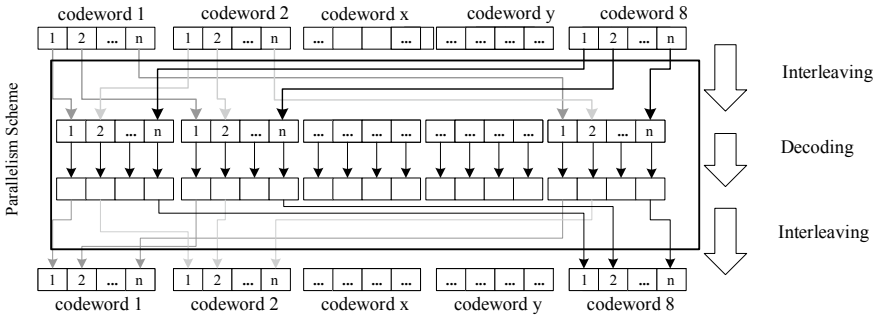


Fig. 3. The multi-codeword parallelism scheme

As shown in Fig. 3, this scheme can decode 8 codeword frames every time. Data elements at the same position from different frames are packed into one 256-bit vector, which enables the processor to achieve a 100% usage rate. Moreover, every codeword has the same parity-check matrix **H**, which means different codeword frames share the same data structure illustrated above. After the decoding process, a similar process has to be performed to unpack results.

4 Experiments and Results

To evaluate the performance of the implementation of QC-LDPC proposed in this paper, some experiments are performed on the RaPro prototyping platform developed by our team in [4].

Fig. 4. The GPP-based 5G RaPro prototyping platform

Figure 4 presents an overall design of this prototyping platform, which is a 16×6 massive MIMO system. The data is received by 16 BS antennas and sent to a server with GPP through optical fibers. To meet the real-time requirements of the RaPro prototyping platform, the throughputs of the encoder and decoder are tested on a processor with Intel Xeon CPU E7-4850 of 2.1 GHz, 2 GB 400 MHz DDR memory.

Table 2. The throughput (Mbps) of the encoder and decoder

Encoder	Decoder		
	Floating point	8-bit fixed point	5-bit fixed point
128.0	21.3	102.3	193.7

The throughputs of the encoder and decoder are given in Table 2. Three different representations of the message between CNs and VNs are compared, and

the result shows that SIMD units can process fixed point more efficiently than floating point. The real-time requirements of the RaPro can be easily satisfied with multi-threading technology if both the encoder and decoder can achieve over 100 Mbps throughput. To test the quantization error of the fixed-point representation, the performance in terms of bit error rate is compared, where BPSK modulation is adopted.

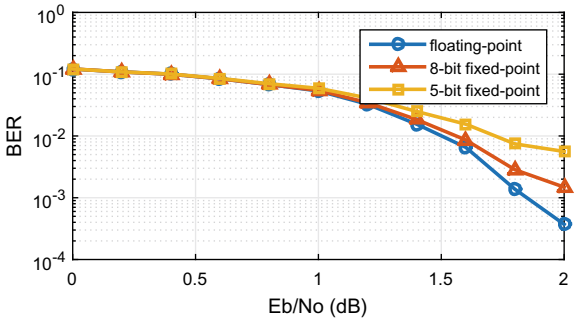


Fig. 5. The quantization error of the fixed-point representation

As shown in Fig. 5, a fixed-point representation instead of floating-point one has a negligible impact in terms of decoding performance. Therefore, to achieve the required throughput, an 8-bit fixed-point representation is selected in the RaPro platform. The BER performance of different modulation schemes in an indoor test on the RaPro platform is given in Fig. 6, where the decoder performs 10 iterations.

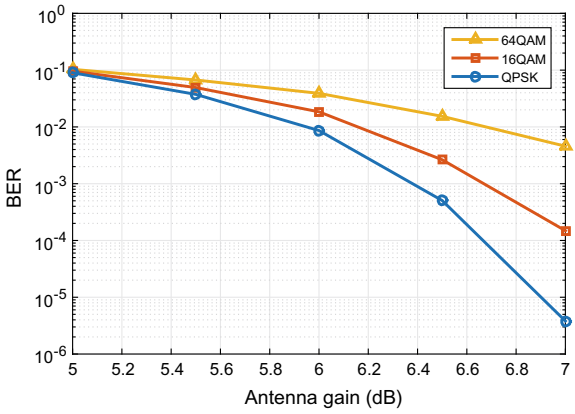


Fig. 6. The BER performances of different modulation schemes

To reach 10^{-3} BER value, the required antenna gain is lower than 10 dB, which proves the practical use and real-time transmission ability of the implemented QC-LDPC on RaPro.

5 Conclusion

In this paper, we applied the QC-LDPC codes adopted in 5G NR on multi-core GPPs of a 5G rapid prototyping system RaPro. Both encoder and decoder can achieve over 100 Mbps throughput with Intel SIMD instructions set, which indicates the real-time processing ability of the proposed software-based implementation. Some scenario tests are performed under the RaPro prototyping testbed, and the results suggest the feasibility and flexibility of GPP-based QC-LDPC implementation.

References

1. Gallager, R.G.: Low density parity check codes. IRE Trans. Inf. Theory **8**(1), 21–28 (1962)
2. Studer, C., Preyss, N., Roth, C., Burg, A.: Configurable high throughput decoder architecture for quasi-cyclic LDPC codes. In: Proceedings of 42nd Asilomar Conference on Signals, Systems and Computers, pp. 1137–1142 (2008)
3. Andrade, J., Falcao, G., Silva, V., Sousa, L.: A survey on programmable LDPC decoders. IEEE Access **4**, 6704–6718 (2016)
4. Yang, X., Huang, Z., Han, B.: RaPro: a novel 5G rapid prototyping system architecture. IEEE Wirele. Commun. Lett. **6**(3), 362–365 (2017)
5. Li, Z., Chen, L., Zeng, L., Lin, S., Fong, W.H.: Efficient encoding of quasi-cyclic low-density parity-check codes. IEEE Trans. Commun. **54**(1), 71–81 (2016)
6. 3GPP TS 38.212. NR; Multiplexing and channel coding, V1.2.0, Release 15 (2017)
7. Fossorier, M.P.C., Mihaljevic, M., Imai, H.: Reduced complexity iterative decoding of low-density parity check codes based on belief propagation. IEEE Trans. Commun. 47(5), 673–680 (1999)
8. Falcao, G., Sousa, L., Silva, V.: Massive parallel LDPC decoding on GPU. In: Proceedings of 13th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP 2008), pp. 83–90 (2008)
9. Jiang, B., Bao, J., Xu, X.: Efficient simulation of QC LDPC decoding on GPU platform by CUDA. In: Proceedings of IEEE International Conference on Wireless Communications and Signal Process, (WCSP), p. 15 (2012)
10. Le Gal, B., Jego, C.: High-throughput multi-core LDPC decoders based on x86 processor. IEEE Trans. Parallel Distrib. Syst. **27**(5), 1373–1386 (2015)

Research on Polarization Modulated QKD Through Overhead Transmission Lines

Yize Tang[1(✉)], Dong He[1], Ying Jiang[1], Qiuhan Wu[1], Tongfei Xia[2], Kejun Xie[2], and Zhiyuan Ye[2]

[1] State Grid Zhejiang Electric Power Co., Ltd. Information & Telecommunication Branch, Hangzhou Zhejiang, 310007 Hangzhou, China
tangyize_sg@126.com

[2] Anhui Jiyuan Software Co. Ltd., SGITG, Hefei 230088, Anhui, China

Abstract. As the intelligence level of power grid growing, the power system relies more on information and communication infrastructure. As a result, the reliability and security of the electric power system have been exposed to threats like illegal hacking interception, tamper, etc., from the informatics field. Quantum Key Distribution (QKD) is considered as a promising technology for power communication network, which can realize One Time Pad (OTP) protocol and prevent key from eavesdropping by exploring the principles of quantum mechanics. In this paper, we study the nonideal characteristics of quantum channel when single photon transmitting over overhead transmission lines. We first present a field test system for BB84 QKD protocol based on polarization modulation, and reveal the relationship between nonideal characteristics of overhead transmission, such as attenuation, internal stress, and strong electromagnetic interference, and received single photons random polarization variation, and then address the above issues by the polarization compensation scheme.

Keywords: Quantum key distribution · Overhead transmission lines · Polarization mode dispersion

1 Introduction

Smart grid is a typical paradigm of cyber-physical system, and there are very strict requirements on the secure communication of power system. Although the secure tunneling, encryption, and authentication based on asymmetric cryptosystem are widely used, it is still possible that the fiber channels could be eavesdropped or confidential message could be decrypted [1]. It is urgent to introduce some new secure communication technology into electric power system communication infrastructure.

Recently, Quantum Key Distribution (QKD) has been the most active research topic in the field of quantum communication [2–5]. Brassard first proposed the QKD protocol based on polarization or phase modulation of single

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 116–124, 2019.
https://doi.org/10.1007/978-981-13-6264-4_14

photon through the aerial space or optical fiber to realize secure key distribution between communication parties [6]. Due to the Heisenberg uncertainty principle and no-cloning principle, QKD can prevent secret key form eavesdropping theoretically, which realizes one-pad system and guarantees so-called unconditional security [7]. The commercialization of QKD has gained more and more attention from academia and industry [8]. In 2016, SGCC applied quantum cryptography to its secure power supply system during G20 Summit in Hangzhou.

The introduction of QKD into the modern power system can dramatically improve ability of resisting information security risks. However, the maturity of QKD protocol realization, rigorous information security requirements and field environment of power system are crucial [9]. The power communication network mainly consists of mass lone-distance overhead lines, and thus we summarize characteristics of quantum channel interference model in following aspects: 1) the attenuation of long-distance transmission of light quantum is relatively high; 2) the negative effects on overhead lines caused by internal stress and temperature variation are non-negligible. The nonidealities of quantum channel may cause unexpected bit flip or phase error for qubits. The modeling and simulation of the above quantum channel is key to QKD for overhead transmission lines.

2 Implementation of Experimental System

In this section, we establish a practical QKD system and measure QBER and QKR to reveal the nonideal characteristics in quantum channel of overhead transmission lines, as shown in Fig. 1. Besides, we also present how to design narrow pulse generation circuit, single photon preparation, and detection. In the end, we propose a method to test the variation of QBER and QKR in the above environment and suggestions on the polarization compensation module to achieve a tradeoff between insertion loss and compensation speed.

2.1 Nonideal Characteristics and Performance Metrics of Polarization Modulated QKD Through Overhead Transmission Lines

Birefringence and dispersion are two unexpected phenomena in single photon propagation. The birefringence effect refers to the fact that when incident light transmit through anisotropic crystal, it is decomposed into two refracted lights whose polarization states are orthogonal, namely, polarization mode dispersion (PMD). Dispersion refers to a physical effect that since the propagation speeds of different components of light are not strictly identical, light pulse can widen when it transmits in the fiber. Although the birefringence and dispersion effects of the single-mode fiber used in the modern day communication system are very weak, the single photon detection accuracy deterioration caused by them at Bob side cannot be negligible.

The performance metrics of QKD system mainly includes quantum key rate (QKR) and quantum bit error rate (QBER). QBER is defined as the ratio of

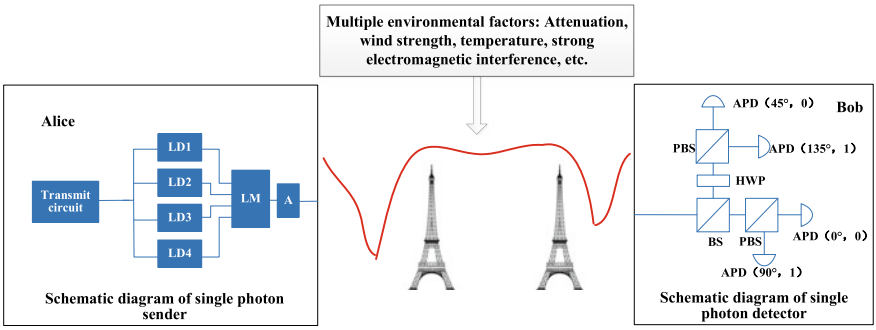


Fig. 1. The system model of quantum key distribution employing BB84 protocol under overhead transmission line environment

number of correct bits in the sifted key to total number of bits in the sifted key, and QKR is defined as the ratio of number of Bob's sifted signal detections to number of Alice's classical signal pulses sent.

Quantum communication equipment and overhead transmission lines are exposed to outdoor environment. Thus, the temperature changes, sweeping wind, vibration, mechanical deformation, and fiber bending can give rise to birefringence, dispersion, and attenuation of light. Furthermore, aforementioned non-ideal characteristics of overhead transmission lines will deteriorate key performance indicators of QKD system, i.e., QKR and QBER.

2.2 Transmit Circuit

As shown in Fig. 2, the function of transmit circuit consists of multiple modules including random number generation (RNG), narrow pulse generation (NPG), delay chips 1, 2, 3, differential D flip-flop (DFF), impedance matching (IM), and amplifier (AM). The main function of transmit circuit is to generated extremely narrow pulse for random single photons preparation at Alices side. Alice first obtains a sequence of random numbers through random number generation (RNG) module, which are input to PC and generates transmission and synchronizing signals through NPG. After delaying and triggering, we can obtain extremely narrow pulse by IM and AM. The extremely narrow pulse is used to stimulate lasing diode to generate light photons.

DC 1 is to ensure that transmission and synchronization signals satisfy time alignment, so as to prevent eve analyze the relative position of the photons in the pulse and learn that the photon is 0 or 1. Meanwhile, we have to adjust the DC 1 to separate synchronization light from transmission light since the synchronous one is stronger, and it is easy to influence the information photons. Accordingly, we will use a similar DC to synchronize them again for the purpose of coherent demodulation of polarization.

DC 2 and DC 3 generate two parallel signals with extremely time difference τ, short-delayed signal is put into the enable pin of DFF, whereas long-delayed

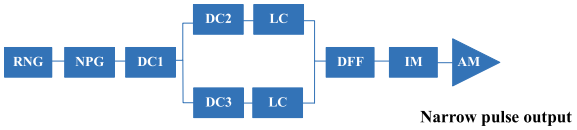


Fig. 2. Schematic diagram of transmit circuit

signal is put into the reset pin of DFF. The D port of DFF keeps a high level. When the clock signal arrives, the output end of DFF will be raised, and when the reset signal arrives, the output will be pulled down again. By doing so, the DFF output is a narrow pulse with a width of τ. The time sequence diagram of narrow pulse generation can be shown in Fig. 3. In order to improve the anti-jamming capability of the signal, we choose differential D flip-flop to ensure that the narrow pulse transmission is not distorted.

IM is to match the impedance of the signal output to the amplifier to a specific value that enables AM to work in the well efficiency. Since the driving voltage of the laser diode is higher than 500 mv, the output signal of the amplifier can generate a light pulse with a width of 1 ns as long as it lasts 1 ns over 500 mv.

2.3 Emission and Detection of Light Photons

The LD shown in Fig. 1 is actually composed of a laser diode and a polarizer, and the laser diode produce classical light pulse which consists of millions of photons, while the polarizer modulates the light pulse to a certain polarization state. The polarized light pulse goes through a two-stage attenuation, and we can obtain the expected single photon to emit into overhead transmission line. When a light pulse contained 0.1 photon, namely every ten pulses have one photon, multiphoton pulses account for less than the 5% of all non-empty pulses.

The receiving light incident on the beam splitter (BS) will be split into a transmit light and a reflect light. In this paper, we used a 50:50 BS, which reflects half of the incident light and transmit remaining light. When the incident light travels through the BS, it may be attenuated, introduced losses depend on polarization. The half wave plate (HWP) and polarizing beam splitter (PBS) make up two measurement bases, i.e., the Horizontal–Vertical Basis and a Diagonal Basis.

Avalanche photodiode (APD) can be used to detect single photon with the avalanche effect. The threshold voltage when avalanche occurs is referred as avalanche voltage. We let the detector work in the Geiger mode, namely it is imposed on a slightly higher voltage than avalanche voltage. The detector is in a critical waiting state. When the photon incident occurs, an avalanche amplification process will occur and then a single photon detection is completed.

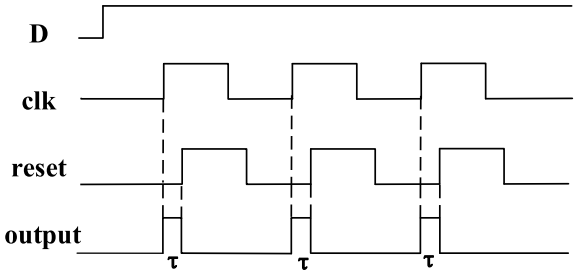


Fig. 3. Time sequence diagram of narrow pulse generation

2.4 The Performance Measurement of Polarization Modulated QKD Through Overhead Transmission Lines

In long-distance overhead transmission lines, the external environment including temperature, stress, and magnetic field results in path loss, dispersion and polarization mode dispersion of photon, which is characterized by a non-unitary transformation. All above nonideal characteristics restrict the transmission efficiency and performance of QKD system.

The loss of fiber can be expressed as attenuation constant α (in decibel per km). If the length of fiber is L then the pass rate of single photon t_f is

$$t_f = 10^{\alpha L/10} \tag{1}$$

When the wavelength is 1330 nm, $\alpha \approx 0.34$ dB/km. And if the wavelength is $\alpha \approx 0.2$ dB/km. Optical fiber path loss and dark count caused by receiver detector noise determine the distance that quantum communication system can achieve.

The polarization change has relatively great influence on the quantum communication, and we will introduce the modeling method with the example of polarization change.

In order to simulate the polarization change in overhead transmission lines, we need to perform comprehensive test of the factors that affect the performance of QKD and gather the polarization change data. We propose to test the polarization Stokes component changes of light photon after traveling through fiber. By doing so, we can evaluate the intensity of polarization change and obtain the correlated data. More specifically, we carry out the test under different environmental and fiber state parameters, and establish the correlation model between each parameter and stokes vector of the polarization state.

In a laboratory simulation environment, we can see from the test results that, after traveling through overhead transmission lines, the change of stokes component S1, S2, and S3 is obvious. Polarization change is frequently and the duration is short. Moreover, the variation of polarization is random and nondirectional.

Apply frequency analysis to the polarization change and calculate the angle of polarization change between each two adjacent sampling points, we can obtain the quantitative intensity of polarization change during the sampling interval.

Theoretically, we can resort to coupled nonlinear Schrodinger equation of nonlinear transmission of optical pulse in birefringent fiber to analyze the polarization change. The amplitudes of light pulse along fast and slow axis are satisfied by

$$i\frac{\partial A_x}{\partial z} + ib' - \frac{1}{2}\beta''\frac{\partial^2 A_x}{\partial t^2} + \gamma\left(|A_x|^2 + \frac{2}{3}|A_y|^2\right)A_x + i\frac{\alpha}{2}A_x = 0, \tag{2}$$

$$i\frac{\partial A_y}{\partial z} + ib' - \frac{1}{2}\beta''\frac{\partial^2 A_y}{\partial t^2} + \gamma\left(|A_y|^2 + \frac{2}{3}|A_x|^2\right)A_y + i\frac{\alpha}{2}A_y = 0, \tag{3}$$

where A_x and A_y are amplitudes of polarized components of light, $b' = ({\beta'}_x - {\beta'}_y)/2$ is the first order group velocity difference between the fast and slow axis, β'' is second order group velocity dispersion, γ is optical fiber nonlinear coefficient, and α is fiber loss.

As for fiber with PMD, denote its PMD coefficient as D_{PMD}, we can divide the fiber into smaller segments, the length of each segment is z_h, and birefringence is Δn, we have

$$D_{PMD} = \sqrt{\frac{8}{3}}\frac{\Delta n}{c}\sqrt{z_h}. \tag{4}$$

We process the light pulse in each segment with coupled wave equation. Between two segments, the orthogonal dominate polarization state experiences a random rotation with angle θ and a random phase delay ϕ.

$$\begin{pmatrix} {A'}_x \\ {A'}_y \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta e^{i\phi} \\ -\sin\theta e^{-i\phi} & \cos\theta \end{pmatrix}\begin{pmatrix} A_x \\ A_y \end{pmatrix}, \tag{5}$$

where θ and ϕ are uniform distributed in $[2, 2\pi]$.

Based on the proposed methods, the correlation model of multiple parameters and quantum communication devices is established accurately under the practical test results. According to the established correlation model, we can process the data with different characteristics, until we find an accurate prediction model. As a result, we can realize the prediction on status of optical fiber so as to optimize the design of the system architecture.

3 Implementation of Experimental System

The entire measurement was carried out with sufficient times. The transmit circuit output narrow pulse with a rate of 4MHz. We measure and observe the QBER and QKR, while the length of fiber varied. The mean photon number was set to be 0.1. The dark count probability of APD we have chosen is 10^{-3}.

Figure 4 shows the QKR versus optical fiber length. As can be seen from Fig. 4, the QKR decreases as the optical fiber length grows. The longer optical fiber is the greater attenuation of optical fiber is. As a result, lower number of light photons arrives at Bob, which causes decrease in QKR.

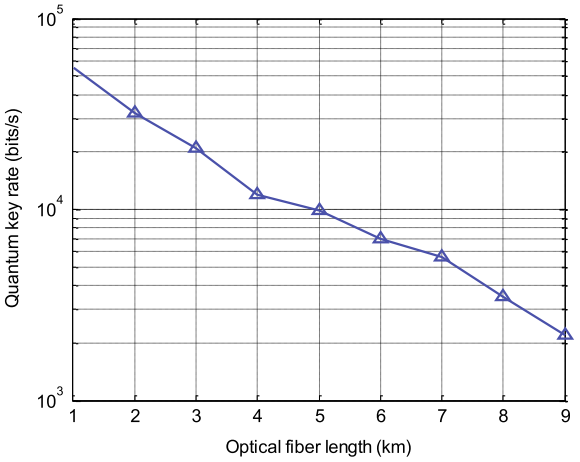


Fig. 4. QKR versus optical fiber length

Figure 5 depicts that QBER increases when optical fiber length grows. The longer optical fiber length is, the intensity of polarization change increases. This inevitably increases detection error.

In the real measurement, we found that the polarization change of the overhead transmission lines caused by the wind was only 0.01 rad/s, where we can use the PZT fiber extruder realizes real-time polarization tracking. Its response time can reach 10 microsecond. When optical fiber across the high-speed railway, it suffers from mechanical vibrations. It is recommended to use a faster electric light plate to compensate for these polarization changes, whose response time

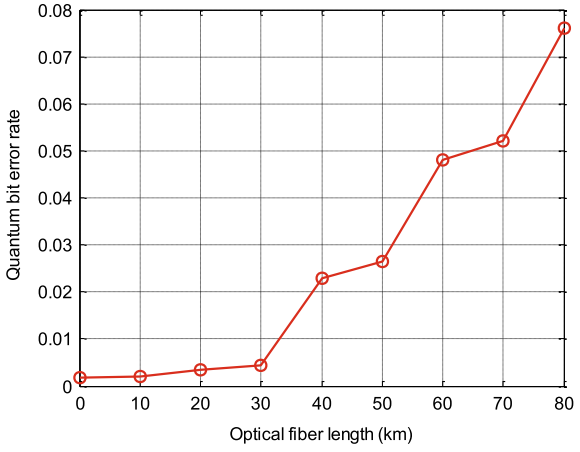


Fig. 5. QBER versus optical fiber length

can be shortened to 10 ns, and polarization tracking speed can be up to a staggering 100 rad/s. To conclude, for overhead transmission lines, we recommend the use of high-speed electric light polarization modulator to ensure the real-time polarization compensation of QKD system. For underground or short-distance fiber, PZT fiber squeezer polarization modulator is more appropriate, since their response speed is moderate, and insertion loss and the cost is much lower than the electro-optic modulator.

4 Conclusion

In this paper, we demonstrate a practical QKD system and measure QBER and QKR to reveal the nonideal characteristics in quantum channel of overhead transmission lines. We present how to design the transmit circuit to produce extremely narrow light pulse. Then we discuss the preparation and detection of the single photon. Then, we propose a test method of QBER and QKR variation under transmission lines environment. Based on experimental results, we provide some constructive suggestion on the selection of polarization compensation module to achieve a tradeoff between insertion loss and compensation speed.

Acknowledgments. This work was supported by State Grid Science and Technology Project "Research and application of quantum communication in electric power communication network" (No. 5211XT160009).

References

1. Shaneman, K., Gray, S.: Optical network security: technical analysis of fiber tapping mechanisms and methods for detection and prevention. In: Proceedings of IEEE Military Communications Conference, pp. 711–716. IEEE Press, New York (2004)
2. Karinou, F., Brunner, H.H., Fung, C.F., Comandar, L.C., Bettelli, S., Hillerkuss, D., Kuschnerov, M., Mikroulis, S., Wang, D., Xie, C., Peev, M., Poppe, A.: Toward the integration of CV quantum key distribution in deployed optical networks. IEEE Photonics Technol. Lett. **30**(7), 650–653 (2018)
3. Mu, Y., Seberry, J., Zheng, Y.: Shared cryptographic bits via quantized quadrature phase amplitudes of light. Opt. Commun. **123**(1–3), 344–352 (1996)
4. Branciard, C., Gisin, N., Kraus, B., Scarani, V.: Security of two quantum cryptography protocols using the same four qubit states. Phys. Rev. A **72**(3), 032301 (2005)
5. Trinh, P.V., Pham, T.V., Dang, N.T., Nguyen, H.V., Ng, S.X., Pham, A.T.: Design and security analysis of quantum key distribution protocol over free-space optics using dual-threshold direct-detection receiver. IEEE Access **6**, 4159–4175 (2018)
6. Bennett, C.H., Brassard, G.: Quantum cryptography: public key distribution and coin tossing. In: Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, pp. 175C–179. IEEE Press, New York (1984)
7. Shor, P., Preskill, J.: Simple proof of security of the BB84 quantum key distribution protocol. Phys. Rev. Lett. **85**(2), 441–444 (2000)

8. Khan, M.M., Murphy, M., Beige, A.: High error-rate quantum key distribution for long-distance communication. New J. Phys. **11**(6), 063043 (2009)
9. Fuloria, S., Anderson, R., McGrath, K., Hansen, K., Alvarez, F.: The protection of substation communications. In: Proceedings of SCADA Security Scientific Symposium, pp. 1–13. Digital Bond Press, Sunrise (2010)

Tunable Power-Dividing Ratio Divider Operation at 3.5 GHz for Future 5G Applications

Menghan Xia, Xiuzhen Luan(✉), Kaizhi Wang, and Kejun Tan

School of Information Science & Technology, Dalian Maritime University, Dalian 116026, China
xiuzhenluan@dlmu.edu.cn

Abstract. A 4-way power divider with continuously tunable power-dividing ratio is presented in this paper. The power divider works in the 3.5 GHz band for 5G applications. The 4-way divider is constituted by connecting three 2-way tunable power dividers, and its power-dividing ratio can be continuously adjusted within a certain range by simply adjusting the bias voltages of three varactors. In this paper, four typical feed requirements of 4-element antenna array are taken, for example, to verify the performances of the 4-way power divider, they are uniform feed of 1: 1: 1: 1; symmetric and non-integer feed of 1: 5.9: 5.9: 1; symmetric and large power-dividing ratio feed of 1: 9: 9: 1; asymmetric non-integer feed of 1: 2.2: 3.8: 7.9. The simulated results show that the radiation patterns of a 4-element microstrip antenna array using this feed network are different at the four different states, and the beamforming and pattern reconfiguration of the antenna array can be realized easily.

Keywords: Tunable power divider · Power-dividing ratio · Antenna array · Beamforming · Pattern reconfiguration

1 Introduction

Current wireless communication technology will not be able to meet the needs of users in the future, this is due to the growth of mobile phone users brought about the exponential growth trend of data services. The solution is known as the fifth generation, or 5G, wireless communication technology [1]. Beamforming is one of the key technologies in 5G mobile communication. Beamforming of the antenna array is mainly realized by controlling the excitation distribution of the antenna array, this can be achieved by designing a proper power-dividing feed network. For different beamforming requirements, the required power-dividing ratios of feed network are different, they may be all sorts of arbitrary power-dividing ratios, such as a large power-dividing ratio, or an asymmetric non-integer power-dividing ratio, etc., this makes it difficult to design the power-dividing feed network. In addition, the power-dividing ratio of a power-dividing feed network is usually fixed, therefore, the corresponding antenna array has a single and fixed radiation pattern. However, in practical applications, it is often necessary to change the radiation pattern of antenna according to the actual

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 125–133, 2019.
https://doi.org/10.1007/978-981-13-6264-4_15

situation, so pattern reconfigurable antenna emerges. The feed network designed with tunable power divider can control the excitation distribution of antenna array element in a certain range, so as to reconstruct the pattern of antenna array, therefore, it is significant to research on the design method of tunable power divider. 3.5 GHz frequency band is one of the working frequency bands of 5G system planed by the Ministry of Industry and Information Technology of China [2], therefore, this paper research on the design method of power-dividing ratio tunable power divider feed network in 3.5 GHz frequency band and its applications in the beamforming and pattern reconfigurable antenna arrays.

2 Design and Simulation of 4-Way Tunable Power Divider

Power divider is a device which can divide the input signal energy into N-way equal or unequal output signal energy, has been used in the feeding network of antenna array widely [3]. In general, the power-dividing ratio of a 2-way power divider is fixed, can be either a loss or lossless 3-port network device, or a 4-port network device. The 3-port network generally takes the forms of T-junction structure or Wilkinson power divider; the 4-port network uses directional coupler or hybrid network, etc. In reference [4], a tunable 2-way power divider in 1 GHz frequency band was proposed and designed. The 2-way power divider has the advantages of wide power-dividing ratio tunable range, low return loss, and large port isolation, therefore, based on this structure, a 4-way tunable power divider in 3.5 GHz frequency band for the 4-element beamforming or pattern reconfigurable antenna array is designed in this paper.

The principle diagram of the 2-way tunable power divider is shown in Fig. 1. The power divider consists of two transmission line sections with electric lengths of 90° and 270°, respectively, two varactors C_D and biasing circuit. The Port 2 is the input port, the Port 1 and the Port 3 is the output ports, and the Port 4 is the isolation port. The power divider can change the output power-dividing ratio by controlling the DC bias voltages of the varactors and then changing the capacitance values of the varactors.

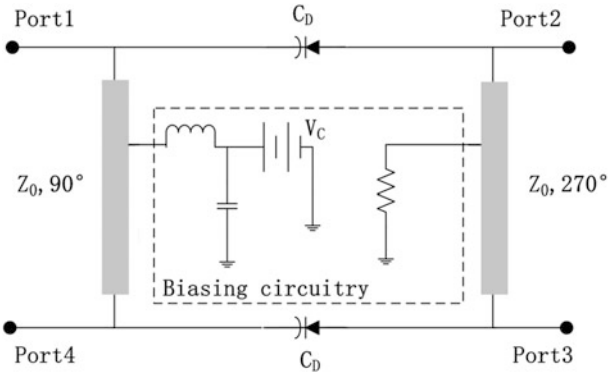


Fig. 1. Principle diagram of the tunable 2-way power divider

The following design formula can be obtained from reference [4].

$$k = \left|\frac{S_{32}}{S_{12}}\right| = \frac{1}{\omega C_D Z_0} \tag{1}$$

In which, k is the power-dividing ratio, C_D is the capacitance value of the varactor. According to the requirements of the operating frequency and the power-dividing ratio range, and the selected Z_0, the appropriate varactor can be selected by using the formula (1), and the sizes of two transmission line sections can be preliminarily determined, then using the simulation software to carry out simulation and optimization, thus the design results that meet the requirements can be achieved. Based on this, a 2-way tunable power divider in 3.5 GHz band is designed in this paper at the first, as shown in Fig. 2. The dielectric substrate with relative permittivity $\varepsilon_r = 4.4$ and thickness h = 0.8 mm is used in the design. In Fig. 2, the characteristic impedances of all transmission line sections are 50 Ω, and the strip width W = 1.54 mm; the lengths of the two transmission line sections are L_1 = 10.34 mm and L_2 = 32.73 mm, respectively; 50 Ω transmission line with length L_P = 6 mm is added at each port; ALPHA's SMV1247 varactors are used in the design. In the bias circuit, C_{blcok} = 10 pF, L_{block} = 80 nH, R_{bias} = 1000 Ω. The power-dividing ratio is controlled by bias voltage V_C. Figure 3 shows the simulated curve of the power-dividing ratio changing with the bias voltages at 3.5 GHz. It can be seen from Fig. 3 that the power-dividing ratio P_3/P_2 of the divider changes from −14.5 dB to 13 dB when V_C changes from 3 V to 6 V, so the tunable range of the power divider is very large.

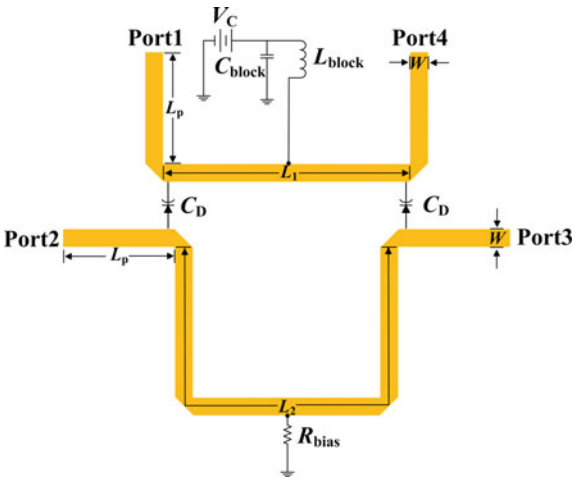


Fig. 2. Configuration of the 2-way tunable power divider

On this basis, three of 2-way dividers are connected to form a 4-way power-dividing ratio tunable power divider, as shown in Fig. 4. Of which, Port 0 is the input port, Port 1, Port 2, Port 3 and Port 4 are output ports. The dimensions of each 2-way

power divider are the same with the dimensions of the power divider shown in Fig. 2, and the power-dividing ratio of 4-way divider is controlled by the bias voltages of the three groups of bias circuits. In order to make bias voltages in three 2-way power divider circuits independent of each other, the isolated DC capacitances C_{DCblock} are added at each place where 2-way power dividers are connected, and takes C_{DCblock} = 250 pF. The loads with resistance R_{L} = 50 Ω are added at the isolated ports of each 2-way power divider.

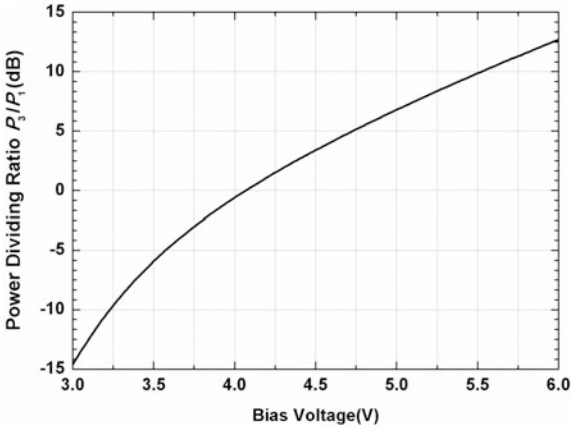


Fig. 3. The power division ratio curve of 2-way tunable power divider with the bias voltage

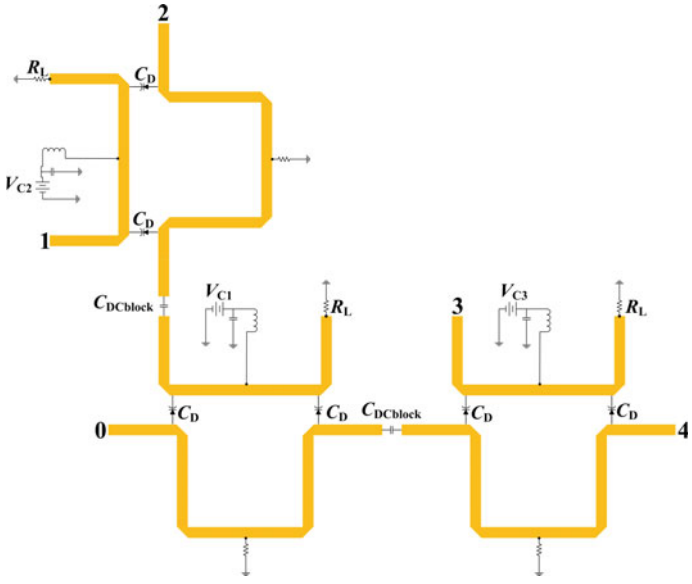


Fig. 4. Configuration of 4-way tunable power divider

3 Applications of the 4-Way Tunable Divider in Beamforming and Radiation Pattern Reconfigurable Antenna Array

Power dividers are widely used in feed networks of array antennas [5]. The excitation amplitude distribution of antenna array element can be changed in a certain range if using the feed network designed by continuously tunable power divider, so the radiation pattern of antenna array can be arbitrary beamforming and reconfigurable. The applications of the 4-way tunable power divider designed in this paper in several typical antenna array feed networks are discussed below.

3.1 Uniform Feed Network

In order to feed for the uniform array, the output power-dividing ratio of the 4-way power divider should be 1 : 1 : 1 : 1, the bias voltages of the varactors are adjusted to be $V_{c1} = V_{c2} = V_{c3} = 4.07$ V. The corresponding simulated results are shown in Fig. 5.

As can be seen from Fig. 5 that $|s_{10}| \approx |s_{20}| \approx |s_{30}| \approx |s_{40}| \approx -7.6$ dB in the band of 3.4 GHz—3.6 GHz, so the power-dividing ratio is about 1 : 1 : 1 : 1.

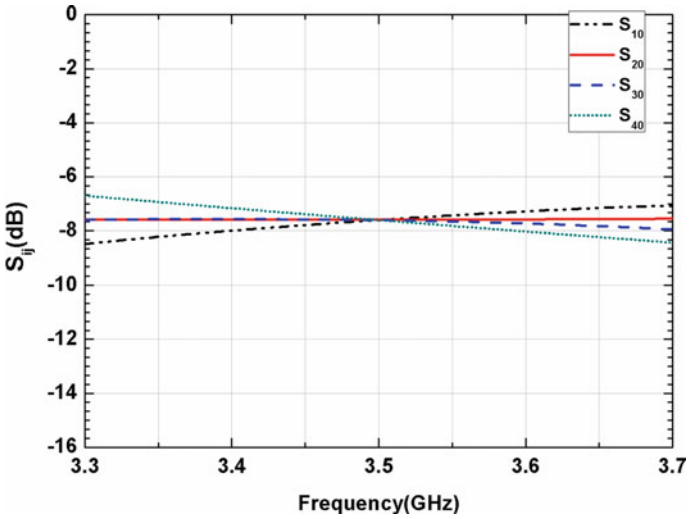


Fig. 5. Simulated results when the power-dividing ratio is 1 : 1 : 1 : 1

3.2 Symmetrical Feed Network with Non-integer Power-Dividing Ratio

A 4-element sector beam antenna array is taken for example. The required power-dividing ratio of the feed network is 1 : 5.9 : 5.9 : 1, this is a symmetric feed network with non-integer power-dividing ratio. According to Fig. 3, the bias voltages of the

varactors are adjusted to be V_{c1} = 4.05 V, V_{c2} = 3.37 V and V_{c3} = 5.15 V. The corresponding simulated results are shown in Fig. 6. It can be seen from Fig. 6 that $|s_{10}| \approx |s_{40}| \approx -5.3$ dB, $|s_{20}| \approx |s_{30}| \approx -12.9$dB in the band of 3.4 GHz—3.6 GHz, so the power-dividing ratio is about 1 : 5.9 : 5.9 : 1.

3.3 Symmetrical Feed Network with Large Power-Dividing Ratio

A 4-element binomial distribution antenna array is taken for example. The required power-dividing ratio of the feed network is 1 : 9 : 9 : 1, this is a symmetric feed network with large power-dividing ratio. According to Fig. 3, the bias voltages of the varactors are adjusted to be V_{c1} = 4.07 V, V_{c2} = 5.44 V, and V_{c3} = 3.26 V. The corresponding simulated results are shown in Fig. 7. It can be seen from Fig. 7 that $|s_{20}| \approx |s_{30}| \approx -5.1$ dB and $|s_{10}| \approx |s_{40}| \approx -14.7$ dB in the band of 3.4 GHz–3.6 GHz, so the power-dividing ratio is about 1 : 9 : 9 : 1.

3.4 Asymmetric Feed Network

Power-dividing ratio of 1 : 2.2 : 3.8 : 7.9 is taken for example. It is an asymmetric feed network with non-integer power-dividing ratio. According to Fig. 3, the bias voltages of the varactors are adjusted to be V_{c1} = 4.80 V, V_{c2} = 4.51 V, and V_{c3} = 4.47 V. The corresponding simulated results are shown in Fig. 8. It can be seen from Fig. 8 that $|s_{10}| \approx -13.3$ dB, $|s_{20}| \approx -9.9$ dB, $|s_{30}| \approx -7.5$dB, and $|s_{40}| \approx -4.4$ dB in the band of 3.4 GHz–3.6 GHz, so the power-dividing ratio is about 1 : 2.2 : 3.8 : 7.9.

Figure 9 shows a 4-element microstrip antenna array fed by the 4-way tunable power divider designed in this paper and its radiation patterns under the above four different bias voltages. The microstrip antenna array element is a rectangular patch

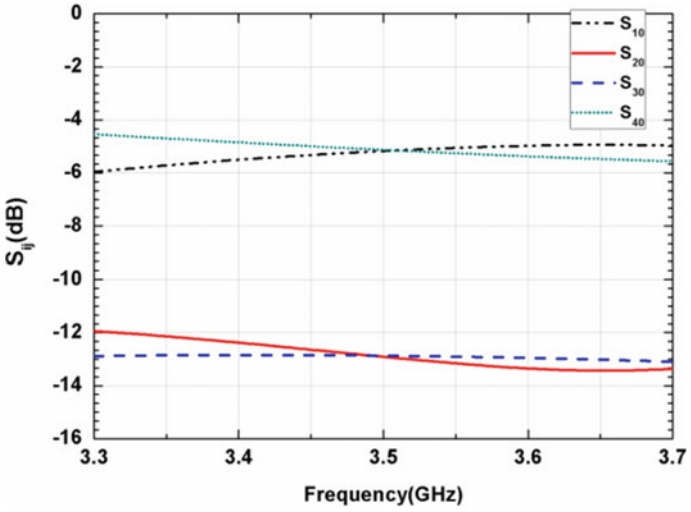


Fig. 6. Simulated results when the power-dividing ratio is 1 : 5.9 : 5.9 : 1

antenna, which is fabricated on a dielectric substrate of relative permittivity $\varepsilon_r = 4.4$. It can be seen from Fig. 9(b) that the radiation patterns of the antenna array are different in the four cases, so the beamforming and pattern reconfiguration of antenna array can be realized by simply changing the bias voltages of the varactors.

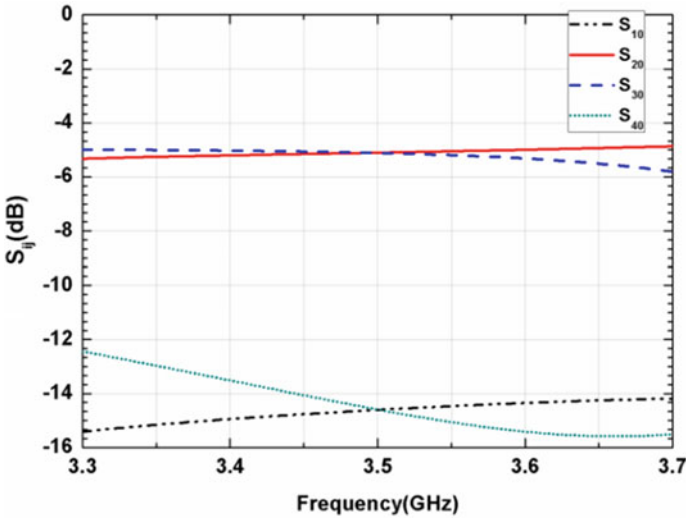


Fig. 7. Simulated results when the power-dividing ratio is 1 : 9 : 9 : 1

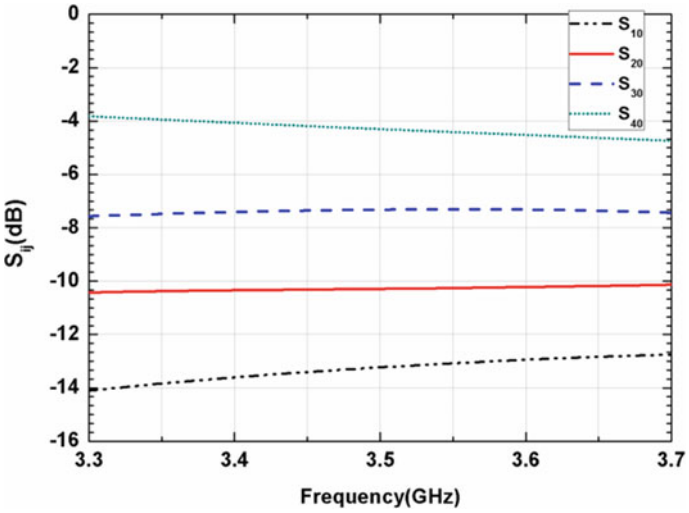
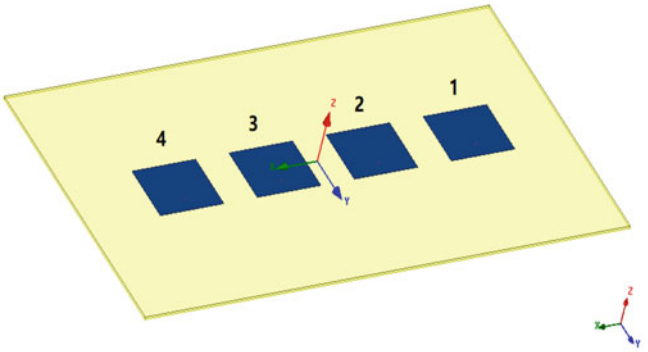
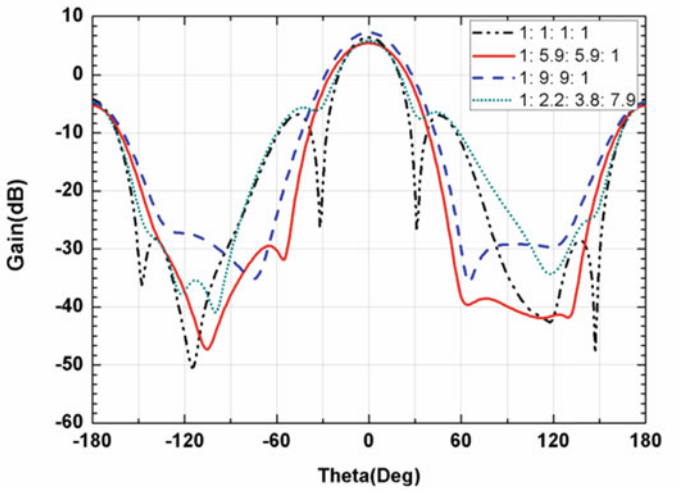


Fig. 8. Simulated results when the power-dividing ratio is 1 : 2.2 : 3.8 : 7.9



(a) 4-element microstrip antenna array



(**b**) radiation patterns

Fig. 9. Microstrip antenna array and its radiation patterns

4 Conclusion

In this paper, a 4-way continuously tunable power divider feed network is designed. The power divider can be used to realize uniform feed, large and non-integer power-dividing ratio feed, asymmetric arbitrary non-integer power-dividing ratio feed, etc., can meet the feed requirements of various antenna arrays within a certain range, can be used to realize the beamforming and pattern reconfiguration by simply adjusting the bias voltages of the varactors in the 4-way divider, so it has a wide application scope.

References

1. Abbas, E.A., Abbosh, A.M.: Tunable millimeter-wave power divider for future 5G cellular networks. In: 2016 IEEE International Symposium on Antennas and Propagation (APSURSI), Fajardo, pp. 1715–1716 (2016)
2. Hong, Wei, et al.: Multi-beam antenna technologies for 5G wireless communications. IEEE Trans. Antennas Propag. **65**(12), 6231–6249 (2017)
3. Cheng, K.K.M., Li, P.W.: A novel power-divider design with unequal power-dividing ratio and simple layout. IEEE Trans. Microw. Theory Tech. **57**(6), 1589–1594 (2009)
4. Cheng, K.K.M., Yeung, S.: A novel rat-race coupler with tunable power dividing ratio, ideal port isolation, and return loss performance. IEEE Trans. Microw. Theory Tech. **61**(1), 55–60 (2013)
5. Al Shamaileh, K., Dib, N., Abushamleh, S.: A dual-band 1:10 wilkinson power divider based on Multi-T-Section characterization of high-impedance transmission lines. IEEE Microw. Wirel. Compon. Lett. **PP**(99), 1–3

Research on Inter-satellite and Satellite-Ground Communication Based on Multi-beam-Phased-Array Antenna

Li Haihao(✉), Huang Yin, Chu Jianxiang, Liu Xiaokuan, Wu Chengchao, and Li Di

Beijing Institute of Remote Sensing Equipment, Beijing, China
lihaihao1982@163.com

Abstract. The low-orbit broadband satellite communication system plays an indispensable role in the civil space-based network and the communication coverage in key areas. It has a very broad application prospect. For the communication system composed of low-orbit satellites, the single satellite coverage is small, and it may be invisible for all the ground gateway stations in some cases. It is the only solution to use inter-satellite link to achieve users' access to ground communication network. By establishing inter-satellite links and making every satellite a node in the communication network, we can form a truly "space-based network". Based on the technology of on-satellite multi-beam-phased-array antenna, the technology of "stare & scan" is adopted in satellite-ground communication to realize coverage of broadband communication in key areas and coverage of narrow-band communication in vast areas. The static routing and dynamic routing is adopted in inter-satellite communication to realize the communication network construction.

Keywords: Low-orbit broadband satellite · Inter-satellite and satellite-ground communication · Phased-array antenna

1 Introduction

In the twenty-first century, technology based on "Internet plus" like big data, cloud services, and smart home has become labels of the era. But there are still 3 billion people far away from broadband Internet in the world. As a result, it will be a huge emerging market to offer satellite Internet for the areas without ground Internet using global satellite coverage. In this background, many international and domestic institutions have put forward the low-orbit broadband satellite communication system, which regards space-based Internet access as the main purpose. It has an excellent application prospect in the extension field of traditional Internet.

Inter-satellite links can achieve connections between satellites. Through on-satellite processing and routing, the number of ground gateway stations can be effectively reduced, which is particularly important for the construction of satellite communication systems that are unable to establish global stations [1]. Inter-satellite links refer to communication links connecting adjacent satellites that are composed of laser or

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 134–141, 2019.
https://doi.org/10.1007/978-981-13-6264-4_16

microwave transmitter, receiver and antenna and so on [2]. Constellations can organize network in the space with inter-satellite links. All satellites can be connected together. The establishment of communication links in the system does not depend entirely on the support of ground gateway stations so that coverage of the system can be enhanced effectively. Especially for low-orbit satellites, the single satellite coverage is small, and it may be invisible for all the ground gateway stations in some cases (such as the middle of the ocean and the land that is unable to establish gateway stations and other areas) [3]. As a result, it is almost the only solution to use inter-satellite link to achieve users' access to ground communication network. Only by establishing inter-satellite links and making every satellite a node in the communication network, can we form a truly "space-based network" [4].

Based on domestic and foreign literature, this article describes the construction of the inter-satellite link in some famous domestic and foreign satellite network systems. Based on the technology of on-satellite multi-beam-phased-array antenna, the technology of "stare & scan" is adopted in satellite-ground communication to realize coverage of broadband communication in key areas and coverage of narrow-band communication in vast areas. The "static routing & dynamic routing" is adopted in inter-satellite communication to realize the communication network construction of inter-satellite link.

2 Development Status

In some satellite network systems that has certain commercial value, Iridium system by Motorola, the Teledesic system proposed by Microsoft and McCormick Co., and the Astrolink system located in the middle and high orbits and Spaceway system claimed that they already have or will achieve on-satellite processing and inter-satellite link in the future [5, 3].

The most important feature of Iridium is achieving global communications through links between satellites, which is equivalent to moving cellular mobile phone systems to the space. Iridium communicates with adjacent satellites with inter-satellite link in Ka-band. Each satellite has four inter-satellite links: two satellites in front of and behind it on the same orbit plane and two satellites on both sides of the adjacent plane. This design of the Iridium means excellent satellite visibility and service coverage on both north and south poles. Since the seam satellites are designed according to polar orbits, the satellite on another orbital plane travels in the opposite direction when the satellite runs in the orbital plane in orbit direction. Horizontal inter-satellite links can switch very quickly to cope with Doppler shifts. Therefore, only satellites in the same direction are supported between Iridium inter-satellite links. Teledesic communication satellite plans to provide worldwide two-way broadband telecommunication services, including computer networks, high-speed broadband Internet access, interactive multimedia, and high-quality voice technologies, etc. Teledesic is known as a space optical fiber network based on satellite network and fast packet switching to provide seamless and global coverage. Each satellite establishes a fast packet switching network and communicates with other satellites on the same and adjacent orbital planes through cross-border channels. The Teledesic system uses each satellite as a communication

node to build inter-satellite exchange network and global coverage. It is a truly "global space Internet". However, due to its huge follow-up investment and low investment return rate, development of Teledesic has suffered big constraints.

On medium and high orbits, Astrolink uses on-satellite processing to increase the efficiency of the system and uses on-satellite exchange to enhance the flexibility of the system. Each satellite is a node of the communication network. This is different from "bent pipe relay" forwarding, whose service needs to be forwarded through satellite node. The Spaceway3 system uses on-satellite processing and unique fast packet switching technology. In order to improve damage resistance of satellite communication system and achieve high-speed access to the ground, the Spaceway3 system plans to add inter-satellite links in the next step.

In January 2015, OneWeb, a satellite network company, announced that it will launch 720 satellites to create a global network without dead ends, in order to provide Internet service to billions of people without access to the Internet in the world. Now it has obtained Virgin and Qualcomm investments and submitted frequency data. In June 2017, FCC approved its application of service in the United States using 720 satellites. In 2018, OneWeb requested an addition of 1,260 satellites, bringing the total number of satellites to 1,980. Constellation built by another company, O3b, is located on the mid-Earth orbit about 8000 km above the equator. It uses Ka-band frequency, covering the area between the north and south latitude of 45°. It mainly offers low-cost high-speed Internet access to an emerging market like Asia, Africa, Latin America, the Pacific, and the Middle East that currently do not have full access to the Internet. The system combines large-scale coverage of satellites with the speed of optic fiber networks. It can offer low-cost, high-speed, low-latency Internet and mobile communication service to billions of consumers and business users in nearly 180 countries. 12 satellites have been launched. The launch quality of each satellite is about 700 kg. Each satellite is equipped with 12 servo antennas and 12 Ka-band transponders. The design life is about 10 years. In the new round of low-orbit satellite communication craze, Google, Samsung, and other companies have put forward their own constellation plans successively. Most companies' plans are now integrated into OneWeb plan after combination of respective advantages. Recently, SpaceX proposed Starlink plan to FCC, which consists of 4425 LEO and 7518 VELO. Its LEO constellation selects Ku/Ka-band frequency to facilitate better coverage, while VLEO constellation selects V-band frequency for signal enhancement and targeted service. The gateway stations, user terminals, satellite payloads will use advanced phased-array antenna technology to achieve interoperability. The low-orbit satellite constellation communication system with inter-satellite links can form a truly global coverage. It can reduce dependence on the ground facilities and provide broadband real-time service. This will be an important part of China's future space-based networks. Starting from "10th Five-Year Plan", many domestic research institutes have already conducted research on constellation system design, networking technology and other key technologies of low-orbit satellite constellations and achieved many results.

3 System Composition and Application

The low-orbit satellite communication system proposed in this paper consists of space segment, ground segment, and user segment. The space segment is an on-orbit satellite that carries a multi-beam-phased-array communication payload with on-satellite processing capabilities. The ground segment includes integrated control center, gateway stations, and monitoring stations. User segment includes user stations and user terminals like smartphones. In this system, satellites process radio signals to achieve uplink and downlink between user stations and gateway stations. Satellites have not only processing capabilities but also authentication control functions. Inter-satellite links are used between satellites to establish network and transmit information. Signal processing and exchange that need to be "landed" are carried out at gateway stations on the ground. User stations are divided into fixed user stations and mobile user stations. Fixed user stations are mainly built on the ground. Mobile user stations are mainly carried on mobile carriers such as aircrafts, ships, and automobiles. User stations are between user terminals and gateway stations in the entire low-orbit satellite communication system. Their main functions are satellite tracking and communication, data preprocessing and storage of user terminals, access, and handover control of user terminals, gateway functions (ground-to-space protocol conversion and processing). Gateway station is an important part of the ground segment. It serves as a link between terrestrial communication network and satellite network. It is the main body of communication between terrestrial network users and satellite network users. It carries out ground-to-satellite service information and control information transmission, mobile switching and related management. At the same time, it achieves functions like receiving and sending satellite signal and converting terrestrial protocols. The gateway station manages satellite resources allocated by system and mobile users who belong to this gateway.

4 System Network Architecture

Figure 1 shows the network architecture of the low-orbit satellite communication system based on multi-beam-phased-array antenna. It consists of three parts: space segment, ground segment, and user segment. The space segment is a low-orbit satellite. In the network architecture, it is an information processing center in the space and acts as a relay station. The ground segment consists of gateway stations, satellite monitoring network, integrated control center and so on. In the network architecture, it is the interface between the system and terrestrial public network. At the same time, user stations can access satellite communication system through gateway stations and interconnect with other terrestrial networks. The integrated control center in the ground segment is responsible for maintaining, monitoring and controlling orbital position and attitude of the satellites and managing satellites' ephemeris etc. Gateway stations work with satellite communication payload and are responsible for call processing, switching, and connection to terrestrial communication networks. Functions of modules such as satellite gateway, Mobile Management Entity (MME), Service Gateway (S-GW),

Packet Gateway (P-GW),and Policy and Charging Rules Functional Unit (PCRF) are integrated in gateway stations.

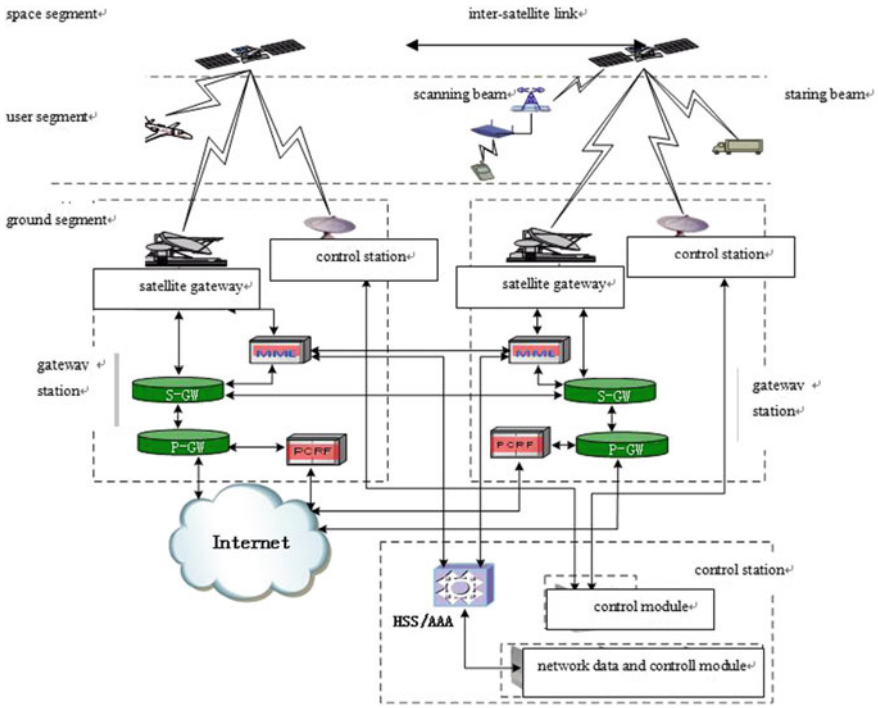


Fig. 1. Low-orbit satellite communication system architecture

5 System Communication Protocols

The overall system communication protocol stack is divided into three layers: physical layer, data link layer, and network and transport layer.

5.1 Interface Protocol Stack

Physical layer: In the upstream direction of gateway station, the upper layer data is mapped to physical data frame and sent using proper modulation and coding. Synchronizing, demodulating, decoding, and channel estimating is achieved in the downstream direction. The acquired data is transmitted up to physical layer. User station processing is similar. The inter-satellite links can use laser or microwave wireless communication. The transmitting end maps data obtained from on-satellite processing to physical data frame and sends it using proper modulation and coding. The processing at the receiving end is opposite of the sending end.

Data link layer: The system process different service according to link state and ensure the quality of service as much as possible and reduce the loss of packet data

effectively. It can also achieve partition and reorganization of resources, retransmission of lost packets and sorting of out-of-order data segments. The upper layer messages processing and IP data packets functions such as header compression, integrity protection, encryption, decryption, packet sorting, and retransmission are achieved. Behavior of the user station in an active connection state, involving establishment, modification, and release of the transmission connection are controlled. Mobility management, satellite selection, and security key control are also achieved.

Network and transport layer: The transport layer uses SCPS-TP protocol to provide reliable end-to-end transmission of commands and data packets and controls transmission rate to avoid congestion and to maintain link bandwidth utilization.

5.2 Satellite-Ground Beam Coverage

Based on on-satellite multi-beam-phased-array antenna technology, in order to reduce the number of phased-array antenna arrays, "stare & scan" technology is adopted in satellite communications to achieve broadband communication coverage in key areas and narrow-band communication coverage in wide areas, shown as Fig. 2. For key areas, based on ephemeris and user location data, the on-satellite phased-array communication payload uses "stare" tracking method to align the beam to the communication target all the time by using a fixed beam and achieve broadband communication in key areas. For large areas, based on a preset beam scanning program, the phased-array communication load on the satellite uses a time-splitting "scan" method to make the beam sweep across communication targets in different areas constantly to achieve narrow-band communication coverage in wide areas.

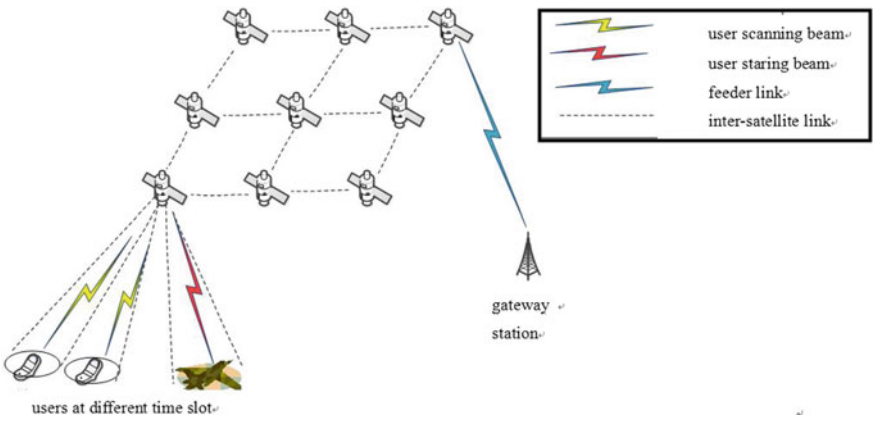


Fig. 2. Satellite-ground beam coverage

In this system, transmitting arrays and receiving arrays of phased-array antenna use sub-array design scheme. The sub-array makes antenna, transmit and receive channels integrate, which meets the need of high integration and high compactness of phased-array antenna. Each receiving and sending sub-array includes 64 channels. The sub-

array uses SPI protocol for external communication, which can realize functions such as single-pass enable control, amplitude-phase control, temperature acquisition and module status acquisition, etc. It mainly consists of three parts: the transmit and receive channel module, the power supply, and control module and cavity. The transmit and receive channel module is implemented by a radio frequency integrated board, which includes multichannel multi-function-integrated SOC chip, RF channel feed network, control, and power feeder network and antenna radiation unit. Power and control modules are integrated power and control boards, including power management circuit and digital control circuit.

5.3 Inter-satellite Routing

When designing satellite networks, the network topology should not be too loose or too dense. The loose network topology will increase hops of the route, thus increasing the transmission delay of the signal. The dense network topology has high requirements on the capacity of inter-satellite link, which increases the difficulty and cost of implementation. In satellite constellations, there are generally four–six satellites around each satellite. In case of most loose topology, the satellite only establishes links with one of the surrounding satellites. In case of most dense topology, the satellite maintains links with multiple surrounding satellites. Two kinds of network topologies are proposed in this paper: (1) mesh topology, that is, each satellite establishes a link with the surrounding four satellites. The partial topology is shown in Fig. 3(a). (2) ring topology, that is, each satellite is only connected to two satellites in front of and behind it. The partial topology is shown in Fig. 3(b).

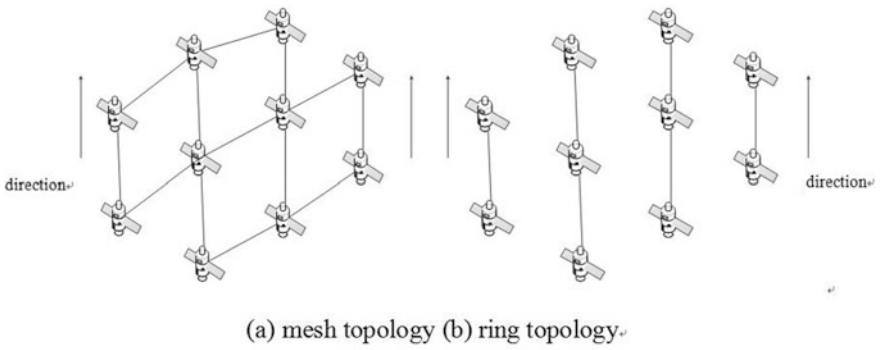


Fig. 3. Topology of inter-satellite networks

As shown in Fig. 2, the inter-satellite link system is divided into upstream access and downstream transmission according to the purpose of information transmission. The purpose of the upstream access is that the user's request and other information are connected to the core network through inter-satellite network, gateway stations, and measurement and control systems. After the information is transmitted into space, the goal is to land as soon as possible. Therefore, in the inter-satellite network, the

destination node selection of upstream information is not unique and can be any satellite that has a feeder link. The purpose of the downstream transmission is that the core network information is passed to the requesting user through measurement and control system, the gateway stations and inter-satellite network. After the information is transmitted into space, there is a clear satellite covering users, and its destination node is unique. Therefore, according to the different types of information transmission, targeted routing protocols are proposed in this paper. In the process of upstream access, the destination node is not unique. Therefore, the dynamic route path-finding is proposed. The routing scheme for downstream transmission is proposed as static route, namely, the improved OSPF routing protocol.

6 Summary and Outlook

In this paper, by constructing the inter-satellite links, each satellite becomes a node in the communication network with "static routing & dynamic routing" technology to form a truly "space-based network". Based on the technology of on-satellite multi-beam-phased-array antenna, the technology of "stare & scan" is adopted in satellite-ground communication to realize coverage of broadband communication in key areas and coverage of narrow-band communication in vast areas.

References

1. Patterson, D.P.: Teledesic: a global broadband network. In: IEEE Aerospace Applications Conference Proceedings, Snowmass at Aspen, CO, USA. IEEE Press, pp. 547–552 (1998)
2. Kota, S.L.: Broadband satellite networks: trends and challenges. In: Wireless Communications and Networking Conference, New Orleans, LA, USA. IEEE Press, pp. 1472–1478 (2005)
3. Zahariadis, T.B., Vaxevanakis, K.G., Tsantilas, C.P., et al.: Global roaming in next-generation networks. IEEE Commun. Mag. **40**(2), 145–151 (2002)
4. Lloyd, W.: Internetworking with Satellite Constellations. University of Surrey (2001)
5. Kota, S.L.: Broadband satellite networks: trends and challenges. Wireless Communications and Networking Conference, New Orleans, LA, USA. IEEE Press, pp. 1472–1478 (2005)

On the Downlink Ergodic Throughput Capacity of Hybrid Wireless Networks with Massive MIMO

Baoju Zhang(✉) and Yipeng Tian

Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China
wdxyzbj@163.com

Abstract. In this paper, the entire network model is a hybrid wireless network model in which each base station is connected to each other via a wired link. In the downlink transmission phase, a large number of antennas (Massive MIMO) are placed at the base station to serve the single node. We derived the ergodic capacity expressions of low SNR and high SNR over Nakagami-m fading, respectively, and analyzed the ergodic capacity change with the number of antennas M on the base station and the shape parameter m.

Keywords: Hybrid wireless network · Massive MIMO · Ergodic capacity · Infrastructure · Fading

1 Introduction and Motivation

Research on large-scale wireless network capacity is initiated by Gupta and Kumar, and they give the definition of network capacity, the network model, and the corresponding research methods [1]. As research continues to deepen, Dousse considered the pros and cons of wireless self-organizing networks and infrastructure networks, he proposed a hybrid wireless network consisting of the above two networks [2]. In [3], to effectively overcome the fast fading impairments, limit the intra-cell interference, and achieve the maximum throughput capacity, Wang introduced the optimal multiple access technique and the successive interference cancelation (SIC) strategy. However, in the downlink transmission phase, just a simple SISO transfer mode, those two technologies are not available in the downlink transmission phase. Zhang found the problem with the downlink transmission in [3], she placed a Point-to-Point MIMO system on the base station to serve the node [4]. But this is not the best solution, Marzetta first initiated research on massive MIMO, massive MIMO has too many advantages over point-to-point MIMO, and he found the small-scale fading can be neglected when the number of antennas tends to infinity [5]. Based on the analysis of [4, 5], massive MIMO can fully address the shortcomings of Point-to-Point MIMO, so we replace the Point-to-Point MIMO on the base station with Massive MIMO to analyze the ergodic capacity over Nakagami-m fading.

The rest of the article is arranged as follows. In Sect. 2, we introduce the established hybrid wireless network model. In Sect. 3, we derive the capacity expression for

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 142–148, 2019.
https://doi.org/10.1007/978-981-13-6264-4_17

the downlink service on the single node with Massive MIMO. In Sect. 4, we derive the downlink ergodic capacity expressions of low SNR and high SNR over Nakagami-m fading with Massive MIMO respectively. In Sect. 5, the whole paper is summarized.

2 Hybrid Wireless Network Model

1. The entire network model consists of n nodes and b base stations.
2. The whole area has size n, each cell has only one base station, and the cell model is a regular hexagon, the parallel side length between the cells is $c = \sqrt{\frac{2\sqrt{3}}{3}\frac{n}{b}}$.
3. Base stations in each cell are connected via a wired infrastructure.
4. In [6], in order to obtain the gains provided by the infrastructure, the number of base stations b must be met $b = \mathrm{O}\left(\frac{n}{\log n}\right)$.

The entire network model is shown in Fig. 1 below.

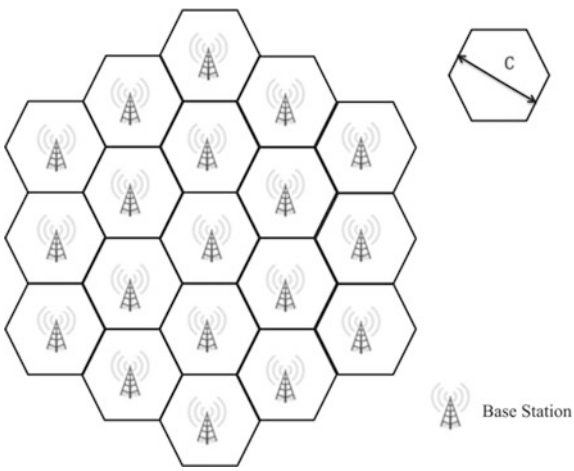


Fig. 1. Hexagon cell with distance c between parallel sides

3 Downlink Ergodic Capacity with Massive MIMO

In [7], the base station with Massive MIMO serve K nodes at the same time.

In the downlink transmission phase, the transmission model can be expressed as

$$x_d = \sqrt{\rho_d} G s_d + N_d \tag{1}$$

$$G = H D_{\beta}^{\frac{1}{2}} \tag{2}$$

x_d is the nodes receives from the base station a $K \times 1$ vector, ρ_d is the quantity which is the proportional to SNR, G is the $K \times M$ propagation matrix in the forward link, it consists of a large-scale fading matrix $D_{\beta}^{\frac{1}{2}}$ and a small-scale fading matrix H, s_d is an $M \times 1$ vector which is of quadrature amplitude modulation (QAM) symbols transmitted by M antennas, N_d is the $K \times 1$ vector of receiver noise whose components are independent and distributed (i.i.d.).

The base station with Massive MIMO just serve single node in the downlink transmission phase, so the transmission model can be expressed as

$$x_{d-1} = \sqrt{\rho_{d-1}}\,\vec{g}\,s_{d-1} + n_{d-1} \tag{3}$$

$\vec{g}$ is the $1 \times M$ propagation vector, and it consists of a small-scale fading vector $\vec{h}$ and a large-scale fading vector $\vec{d_{\beta}^{\frac{1}{2}}}$.

In [5], the channel coefficient between the kth node and the mth antenna of the base station is defined as

$$g_{mk} = h_{mk}\sqrt{\beta_k} \tag{4}$$

h_{mk} is the small-scale fading coefficient, and β_k is the large-scale fading coefficient. In [8], we can use the following expression to represent β_k

$$\beta_k = \varphi d_k^{-\alpha}\zeta_k \tag{5}$$

The φ is a constant related to the antenna gain and carrier frequency, d_k is the distance between the base station and kth node, α is the path loss exponent, ζ_k is the log-normal shadowing with $10\log_{10}\zeta_k \sim N\left(0, \sigma_k^2\right)$.

Combined with the above analysis and introduction, when the downlink transmission service is performed on a single node, assuming that the base station has perfect CSI knowledge. The signal expression after preprocessing through MF is

$$\vec{g}^{H}\,x_{d-1} = \sqrt{\rho_{d-1}}\,\vec{g}^{2}\,s_{d-1} + \vec{g}^{H}\,n_{d-1} \tag{6}$$

H represents conjugate transpose, so the downlink ergodic capacity serving the single node with Massive MIMO is

$$\bar{C} = E\left\{\log_2\left(1 + \rho_d\beta\left\|\vec{h}\right\|^2\right)\right\} \tag{7}$$

$$\left\|\vec{h}\right\|^2 = \sum_{i=1}^{M}|h_m|^2 \tag{8}$$

4 Downlink Ergodic Capacity Over Nakagami-m Fading

The pdf of a Nakagami-m random variable $|h|$ is

$$f(h) = \frac{2}{\Gamma(m)}\left(\frac{m}{\Omega}\right)^m h^{2m-1} e^{\frac{-mh^2}{\Omega}}, h \geq 0 \tag{9}$$

$\Gamma(\cdot)$ is the Gamma function, $\Gamma(m) = \int_0^\infty x^{m-1}e^{-x}dx$, m is the shape parameter, Ω stands for the controlling spread and $\Omega = 1$ for Nakagami-m fading. $|h_m|^2$ satisfies Gamma distribution with $|h_m|^2 \sim \Gamma\left(m, \frac{1}{m}\right)$, so the pdf of $|h_m|^2$ is

$$f(x) = \frac{m^m}{\Gamma(m)} x^{m-1} e^{-mx}, x \geq 0 \tag{10}$$

So, $\left\|\vec{h}\right\|^2 = \sum_{i=1}^{M} |h_m|^2$ satisfies Gamma distribution with $\left\|\vec{h}\right\|^2 \sim \Gamma\left(Mm, \frac{1}{m}\right)$ too, the pdf of $\left\|\vec{h}\right\|^2$ is

$$f(x) = \frac{m^{Mm}}{\Gamma(Mm)} x^{Mm-1} e^{-mx} \tag{11}$$

Combined with the above analysis, downlink ergodic capacity can be expressed as

$$\bar{C} = \int_0^\infty \log_2(1 + x\beta\rho_d) \cdot \frac{m^{Mm}}{\Gamma(Mm)} x^{Mm-1} e^{-mx} dx \tag{12}$$

We analyze ergodic capacity in two scenarios of low SNR and high SNR, respectively. At low SNR, we employ the approximation $\ln(1+x) \approx x$ and we can get

$$\bar{C} = \int_0^\infty \frac{\beta\rho_d}{\ln 2} x \cdot \frac{m^{Mm}}{\Gamma(Mm)} x^{Mm-1} e^{-mx} dx \tag{13}$$

Following the equation $\int_0^\infty x^n e^{-\mu x} dx = n! \cdot \mu^{-n-1}$ we can get

$$\bar{C} = \frac{\beta\rho_d m^{Mm}}{\Gamma(Mm)\ln 2} (Mm)! \cdot m^{-[Mm+1]} \tag{14}$$

Since the gamma function satisfies the equation $\Gamma(n) = (n-1)!$, so we can get the final expression of downlink ergodic capacity over Nakagami-m fading with Massive MIMO at low SNR is

$$\bar{C} = \log_2^e M\varphi d^{-\alpha}\zeta\rho_d \tag{15}$$

It can be seen that the ergodic capacity is not affected by the shape factor m in the low SNR scenario.

Simulation results are shown as follows in Fig. 2.

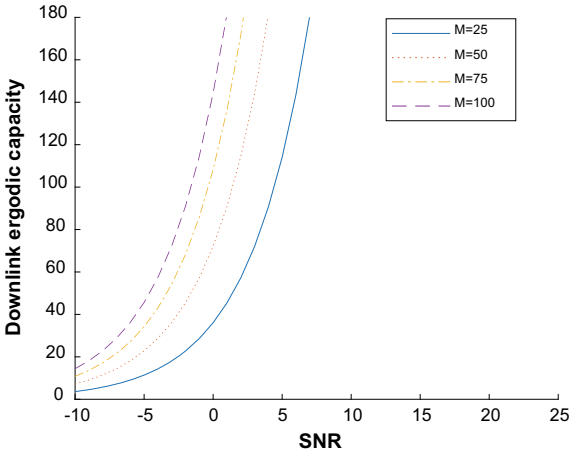


Fig. 2. Ergodic capacity over Nakagami-m fading at low SNR

At high SNR, we use the approximation $\log(1+x) \approx \log(x)$ to calculate the ergodic capacity, so we can get

$$\bar{C} = \int_0^{\infty} \log(\beta\rho_d x) \cdot \frac{m^{Mm}}{\Gamma(Mm)} x^{Mm-1} e^{-mx} dx \tag{16}$$

Simplify the equation deformation to

$$\bar{C} = \log(\beta\rho_d) + \frac{1}{\Gamma(Mm)\ln 2}\left[\int_0^{\infty} \ln x \cdot x^{Mm-1} e^{-x} dx - \ln m \int_0^{\infty} x^{Mm-1} e^{-x} dx\right] \tag{17}$$

Following the equation $\int_0^{\infty} x^n e^{-\mu x} \ln x dx = \frac{n!}{\mu^{n+1}}\left[1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - C - \ln\mu\right]$, $\mu > 0$, we can get the final expression of downlink ergodic capacity over Nakagami-m fading with Massive MIMO at high SNR is

$$\bar{C} = \log\left(\frac{\varphi d^{-\alpha}\zeta\rho_d}{m}\right) + \log_2^e\left[\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{Mm-1}\right) - C\right] \tag{18}$$

$$C = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \ln n\right) \approx 0.57721 \tag{19}$$

When the value of the shape factor m is 1, the ergodic capacity changes with the number of antennas M. The simulation results are shown as follows in Fig. 3.

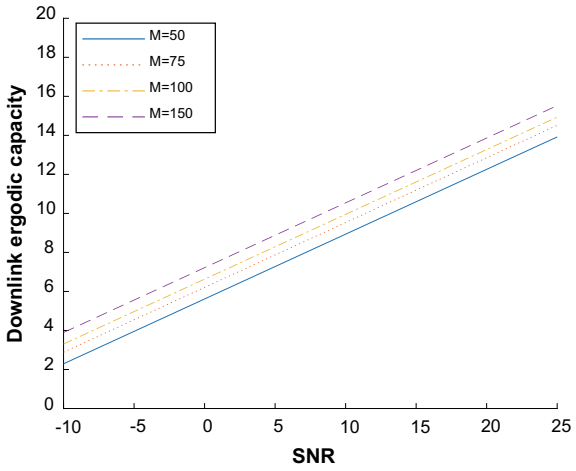


Fig. 3. Ergodic capacity over Nakagami-m fading at high SNR (m = 1)

When we set the M as 100, the comparison ergodic capacity changes with the shape factor m, and we find that the ergodic capacity increases as the shape factor m increases. The simulation results are shown as follows in Fig. 4.

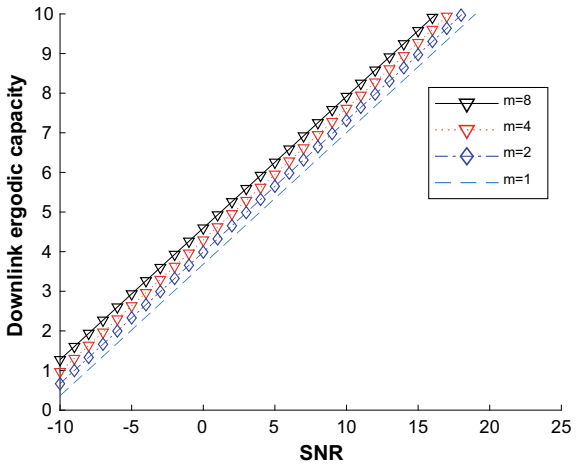


Fig. 4. Ergodic capacity over Nakagami-m fading at high SNR (M = 100)

5 Conclusion

In this paper, we mainly analyzed the downlink ergodic capacity of the hybrid wireless network over Nakagami-m fading with massive MIMO. From the above analysis, we obtained the closed expressions of ergodic capacity (low SNR and high SNR) in the downlink transmission scenario. At low SNR, ergodic capacity is not affected by the shape factor m, we can find that when the number of antennas M at the base station increase, the ergodic capacity increases too. At high SNR, first, we find that when the shape factor m takes a value (m = 1), the ergodic capacity increases with the number of antennas M; second, we find that when M takes a value (M = 100), the ergodic capacity increases with the shape factor m.

References

1. Gupta, P., Kumar, P.R.: The capacity of wireless networks. IEEE Trans. Inf. Theory **46**(2), 388–404 (2000)
2. Dousse, O., Thiran, P., Hasler, M.: Connectivity in ad-hoc and hybrid networks. In: Proceeding of the 2002 INFOCOM, vol. 2, pp. 1079–1088 (2002)
3. Wang, X., Liang, Q.: On the ergodic throughput capacity of hybrid wireless networks over fast fading channels, In: Proceedings of the 2013 IEEE International Conference on Communications, Budapest, Hungary (2013)
4. Zhang, B., Wang, Y., Wang, W., Tian, Y.: On the downlink throughput capacity of hybrid wireless networks with MIMO. IEEE Access **5**, 26086–26091 (2017)
5. Marzetta, T.L.: Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans. Wireless Commun. **9**(11), 3590–3600 (2010)
6. Liu, B., Liu, Z., Towsley, D.: On the capacity of hybrid wireless networks. In: INFOCOM, vol. 2, pp. 1543–1552 (2003)
7. Rusek, F., et al.: Scaling up MIMO: opportunities and challenges with very large arrays. IEEE Signal Process. Mag. **30**(1), 40–60 (2013)
8. Zheng, K., Ou, S., Yin, X.: Massive MIMO channel models: a survey. Int. J. Antennas Propag. **2014**, 10 (2014). Article ID 848071

Delayed Self-homodyne Direct Detection for Constant Envelope OFDM

Yupeng Li[1,2(✉)]

[1] Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China
fx_lyp@163.com
[2] College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin 300387, China

Abstract. A novel delayed self-homodyne direct detection (DSHDD) structure for constant envelope OFDM (CE-OFDM) system was proposed. The data information was loaded on the phase of light and a delayed coherent structure was used to get the phase information. We demonstrated its feasibility and effectiveness with numerical simulations. Both OFDM-QPSK and OFDM-16QAM were evaluated. The results showed that DSHDD CE-OFDM performs better than conventional direct detection (Con-DD) OFDM system in resisting fiber nonlinearity.

Keywords: Constant envelope · OFDM · Self-homodyne · Direct detection · Nonlinearity

1 Introduction

The data traffic in short and medium distance optical transmission, such as metro network and access network, is increasing rapidly [1, 2]. To cope with the rapid growth, cost-effective optical transceivers which have low power consumption, high resilience to fiber impairments, and high spectral efficiency (SE) gain a lot of attention. Single-polarization direct detection has much simpler and economic efficiency optical structure, which may become a suitable solution for metro and access networks.

Orthogonal frequency-division multiplexing (OFDM) [3], which has high spectral efficiency, dynamic spectrum allocation and high tolerance to chromatic dispersion (CD) and polarization mode dispersion (PMD), is suitable to be utilized for short and medium distance transmission. However, due to the inherent high peak-to-average power ratio (PAPR), its performance is severely degraded because of a nonlinear effect introduced during the transmission.

Constant envelope OFDM (CE-OFDM) scheme has been proposed to solve the problem of high PAPR [4]. Phase modulator is used and the better fiber nonlinear tolerance is obtained. However, complex coherent reception is required for CE-OFDM to extract the phase information.

Although coherent reception is an effective way to improve the receive sensitivity and transmission capacity, coherent reception with local oscillator (LO) has a relatively complex and expensive structure. The operation complexity and cost increase

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 149–156, 2019.
https://doi.org/10.1007/978-981-13-6264-4_18

significantly [5, 6]. For the short and medium distance transmission, direct detection is more reasonable.

Self-homodyne has been proposed as an effective way to reduce the cost and complexity of receiver. Several self-homodyne schemes have been proposed [7–10]. One is to use two orthogonal polarizations to send signal and local oscillator (LO), respectively, which needs polarization multiplexed structure and sacrifices half of the spectral efficiency. Furthermore, optical carrier extraction technique, by separating the LO from the signal, is adopted for self-homodyne coherent detection. But the requirement on the narrow-band filter is stringent. Another scheme is based on few-mode fibers (FMF) or multicore fibers (MCF). One of the fiber modes or fiber cores is used to transmit LO, the others are used to transmit optical signal. However, the system is relatively complex. It needs to be pointed out that commercial coherent reception equipment is still required for all the self-homodyne schemes above, which would increase the cost of receiver.

In this paper, we demonstrate a novel delayed self-homodyne direct detection (DSHDD) scheme for CE-OFDM. The optical signal and carrier transmit in turn. A delayed coherent structure and balance photodiode are used for signal reception. The scheme is relatively simple but effective. The performance of fiber nonlinearity tolerance is improved with this scheme. Numerical simulations results show that DSHDD CE-OFDM performs better than conventional direct detection OFDM system with intensity modulator in resisting fiber nonlinearity. Furthermore, it shows good laser linewidth tolerance, which will reduce the requirement of laser in the practical system.

2 Theoretical Analysis

Here, we describe in detail the theory of DSHDD CE-OFDM.

Figure 1(a) shows the schematic of CE-OFDM transmitter. Phase modulator is used for signal modulation. The output of OFDM coder with proper amplitude is loaded to the phase modulator. The output optical signal has constant intensity and the data information is carried by the phase of optical signal. Each electrical OFDM is followed with a null duration. The length of the null duration is equal to an OFDM symbol duration, so the phase of modulated optical light is shown in Fig. 1(c).

It is worth mentioning that constant envelope optical OFDM has low PAPR (0 dB), so it has better resilience to fiber nonlinearity. Furthermore, the optical phase noise introduced during signal transmission turns to be additive, and the phase noise distortion can be avoided by simply leaving the subcarriers around direct current (DC) empty. The complexity of signal processing at the receiver side could be reduced.

Figure 1(b) shows the schematic of the DSHDD receiver. The received optical signal is split into two branches, and one branch is delayed by an OFDM symbol duration, as shown in Fig. 1(d). Then, the two branches are combined and detected by the balanced photodiode (BPD).

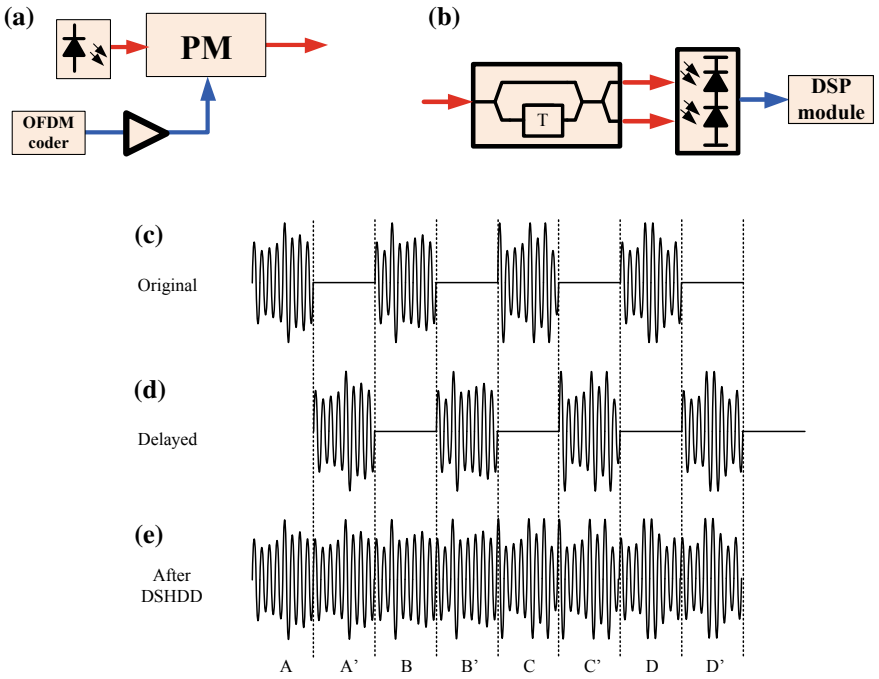


Fig. 1. Schematic of (a) CE-OFDM transmitter and (b) DSHDD receiver. The phase information of (c) original received signal (d) delayed received signal. (e) The electrical intensity signal after detection

We assume the received signal is

$$R(t) = \begin{cases} \sqrt{P}e^{j(2\pi f_c t + \phi_s + \triangle\phi)} & 2kT < t < (2k+1)T \\ \sqrt{P}e^{j(2\pi f_c t)} & (2k+1)T < t < 2(k+1)T \end{cases} \tag{1}$$

where P is the received optical power, f_c is the frequency of optical carrier, T is the OFDM symbol duration, ϕ_s represents the data information, $\triangle\phi$ is the phase difference between consecutive two optical OFDM symbols, which is mainly caused by laser linewidth and phase noise during transmission. The output of the hybrid is

$$\begin{aligned} \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} &= H \cdot \begin{bmatrix} R(t) \\ R(t+T) \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & j \\ j & 1 \end{bmatrix} \cdot \begin{bmatrix} \sqrt{P}e^{j(2\pi f_c t + \phi_s + \triangle\phi)} \\ \sqrt{P}e^{j(2\pi f_c t)} \end{bmatrix} \end{aligned} \tag{2}$$

where E_1 and E_2 are the upper and lower output of hybrid, respectively. H is the transfer function of hybrid.

The output of BPD is

$$\begin{aligned} s(t) &= r(E_1 \cdot E_1^* - E_2 \cdot E_2^*) + n \\ &= rP\sin(\phi_s + \triangle\phi) + n \end{aligned} \tag{3}$$

where r denotes the responsivity of BPD, n is the noise of BPD. The optical phase information is transferred to electrical intensity signal. Then the signal is stored and processed in DSP module to recover the original information.

The output electrical signal of BPD is shown in Fig. 1(e). The consecutive two detected symbols have the same information, as shown in Fig. 1(e), symbols A and A', B, and B', …, have same information. In the following DSP module, symbols A', B'…, is selected to recover the original information.

3 Simulation Setup and Results

Figure 2 shows the simulation setup. The linewidth of the laser is 100 kHz. QPSK and 16QAM are mapped into each subcarrier of OFDM. No clipping are used for the OFDM signals. The sampling rate in the simulation is 10GSa/s.

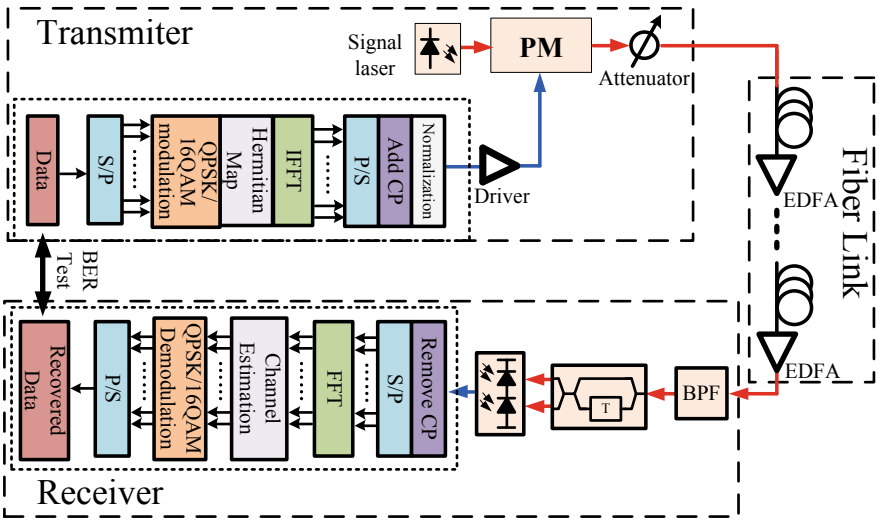


Fig. 2. Simulation setup of single-channel delayed self-homodyne direct detection constant envelope OFDM

At transmitter side, the electrical OFDM signal is generated in MATLAB program. 110 subcarriers are used to carry data. 7 subcarriers around DC are left empty to avoid phase noise distortion. 10 subcarriers around Nyquist frequency are left empty to avoid high frequency distortion. The data vector before IFFT should be Hermitian conjugate

symmetric to get real-valued electrical OFDM signal. The length of final constructed Hermitian vector before IFFT is 256. Cyclic prefix (CP) of 32 samples is added to avoid ISI. After normalization, the electrical OFDM signal is amplified to proper amplitude and loaded to the phase modulator. The followed attenuator is used to control the launch power into fiber link.

Standard single mode fiber (SSMF) is used as fiber link, with parameters as follows: attenuation, dispersion index, nonlinear index and effective core area of fiber are 0.2 dB/km, 16 ps/nm/km, 2.6×10^{-20} m^2/W, and 80×10^{-12} m^2, respectively. The EDFAs are used for signal amplification and adjust OSNR of received signal. The

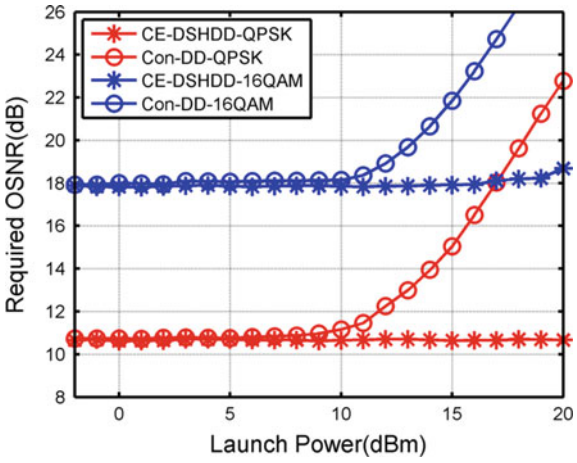


Fig. 3. Required OSNR of 80 km single channel transmission with different launch power into fiber

bandwidth of band-pass filter (BPF) is 50 GHz. The received signal is detected with the proposed structure and processed in the DSP module.

Figure 3 shows the required OSNR for BER = 3.8×10^{-3} after the transmission over 80 km SSMF. The system performance keeps stable with low launch power, which is in the linear region of fiber. CE-DSHDD requires an OSNR of about 10.8 dB and 17.9 dB for QPSK and 16QAM, respectively, which is nearly same as conventional direct detection (Con-DD). With increasing launch power, the nonlinear effect will influence the system performance. For Con-DD, the required OSNR begins to increase when launch power larger than 10 dBm. However, benefit from low PAPR, CE-DSHDD does not suffer obvious nonlinear distortion until the launch power reaches 20 dB.

Figure 4 shows the simulation results with 400 km transmission. The optical signal suffers more severe nonlinear distortion with longer fiber transmission. Simulation result shows CE-DSHDD still has good nonlinear tolerance with launch power less than 10 dBm and 8 dBm for QPSK and 16QAM, while Con-DD suffers severe

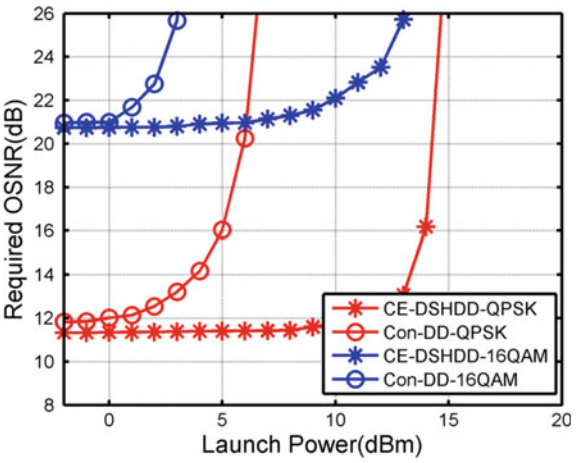


Fig. 4. Required OSNR of 400 km single channel transmission with different launch power into fiber

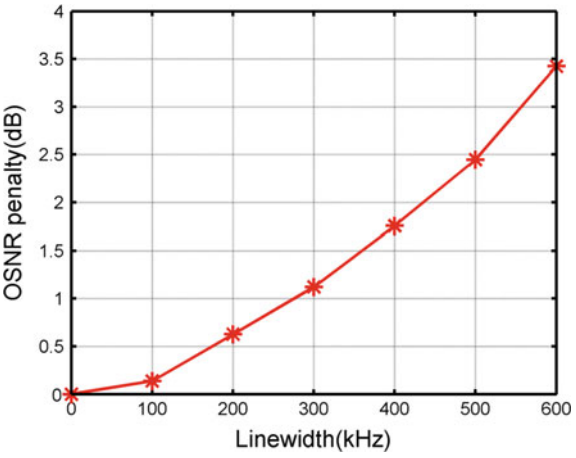


Fig. 5. Laser linewidth tolerance of CE-DSHDD in 80 km transmission

nonlinear distortion and should keep the launch power less than 2 dBm to avoid 1 dB penalty.

The effect of phase noise will influence the system performance of phase modulation, and the phase is mainly caused by the laser linewidth. The linewidth tolerance in 80 km transmission is tested and the result is shown in Fig. 5. We leave seven subcarriers around DC empty to avoid phase noise distortion. The OSNR penalty is about 0.6 dB when the linewidth increase to 200 kHz, and less than 3.5 dB when the linewidth increase to 600 kHz. The proposed CE-DSHDD shows good linewidth tolerance, which is suitable for the practical application.

All the simulation results show that DSHDD CE-OFDM has an obvious advantage in resisting fiber nonlinearity.

Several advantages of DSHDD CE-OFDM are obtained through the analyses above. First, PAPR of CE-OFDM decreases to 0 dB because of its constant amplitude, and hence it has better fiber nonlinearity tolerance. Furthermore, CE-OFDM scheme does not need phase noise compensation, so the signal processing at receiver side can be simplified. The reduced computation complexity is beneficial for the real-time system. Last, DSHDD does not need LO and coherent detection, so the complexity and cost of receiver could be reduced effectively.

4 Conclusion

In summary, a delayed self-homodyne direct detection structure for constant envelope OFDM scheme was demonstrated. Performances of OFDM-QPSK and OFDM-16QAM are investigated. This is a relatively simple but effective scheme. Simulation results show that the proposed scheme performs well in resisting fiber nonlinearity. Furthermore, the proposed structure has good laser linewidth tolerance, so the proposed scheme could become an effective alternative in short and medium distance transmission.

Acknowledgements. This work is supported by the Doctor Fund of Tianjin Normal University (52XB1505).

References

1. Wong, E.: Next-generation broadband access networks and technologies. J. Lightw. Technol. **30**(4), 597–608 (2012)
2. Gaudino, R.: Advantages of coherent detection in reflective PONs. In: Proceeding of the OFC/NFOEC (2013). Paper OM2A.1
3. Armstrong, J.: OFDM for optical communications. J. Lightw. Technol. **27**(3), 189–204 (2009)
4. Thompson, S.C., Ahmed, A.U., et al.: Constant envelope OFDM. IEEE Trans. Commun. **56** (8), 1300–1312 (2008)
5. Zhou, H., Li, X., et al.: Joint timing/frequency offset estimation and correction based on FrFT encoded training symbols for PDM CO-OFDM systems. Opt. Express **24**(25), 28256–28269 (2016)
6. Ouyang, X., Dobre, O.A., et al.: Unbiased channel estimation based on the discrete fresnel transform for CO-OFDM systems. Photon. Technol. Lett. **29**(8), 691–694 (2017)
7. Luís, R.S., Puttnam, B.J., et al.: Self-homodyne detection of polarization-multiplexed pilot tone signals using a polarization diversity coherent receiver. In: Proceedings of the ECOC (2013)

8. Xu, L., Hu, J., et al.: Coherent optical OFDM systems using self optical carrier extraction. In: Proceeding of OFC/NFOEC (2008). Paper OMU4
9. Hu, R., Gui, T., et al.: Delayed self-homodyne detection for OFDM-PON downstream. In: Proceeding of the OFC/NFOEC (2014). Paper W2A.18
10. Chen, Y., Li, J., et al.: Novel MDM-PON scheme utilizing selfhomodyne detection for high-speed/capacity access networks. Opt. Express **23**(25), 32054–32062 (2015)

Calculation and Experimental Analyze of the Oceanic Expendable Instruments' Communication Channel Distributed Parameters

Ning Liu, Chen Xu(✉), and Rubin Zhang

National Ocean Technology Center City, Tianjin 300112, China
liun0726@163.com, ll2ll82898@qq.com, zhrbl233@126.com

Abstract. The purpose of this paper is to present a new method for predicting the key parameters of the communication channel of underwater expendable measuring instruments, which is widely used in ocean exploration, then compare the theoretical and experimental results. This paper mostly presented a method to model the communication channel, and analyzed the key parameters of cable and coils winding, respectively, in theory. To calculate the distributed parameters of cable, the classical method was multiplied with the differential and integral calculus. To calculate the distributed parameters of spool, a double-layer modal for capacitance calculation, which is more concise was proposed according to the actual structure of the solenoid. Furthermore, the values of every parameter had been verified with experimental measurements. The test results are consistent with the theoretical results. It shows that this theoretical analysis is feasible and can provide guidance for relevant theoretical research and applications.

Keywords: Expendable measuring instruments · Time-varying channel · Distributed parameters

1 Introduction

The expendable instruments' main problem of long-distance digital communication in seawater are transmission signal waveform distortion,which caused by the change of the cable characteristic parameters and the influence of the seawater. It is difficult to identify the transmission data at the receiving end. Modeling and analysis of the communication channel's key parameters is a fundamental step for solving this problem.

The purpose of this paper is to present a new method for predicting the key parameters of the communication channel of vessel carried undersea expendable measuring instruments which are widely used in ocean exploration, then compare the theoretical and experimental results. These instruments as shown in Fig. 1 which is disposable is used for measuring the sea environment parameters. The channel is composed with the receiver in which installed 1000 m above coil windings and the detector underwater also included exceed 1000 m coil windings in order to ensure

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 157–170, 2019.
https://doi.org/10.1007/978-981-13-6264-4_19

1000 m depth measurement. The detector was launched by a special device, with the submerged detector, the sensor, and processor in detector complete the data real-time measuring, manipulating, and finally transferring in the form of digital signal along the long-thin transmission line.

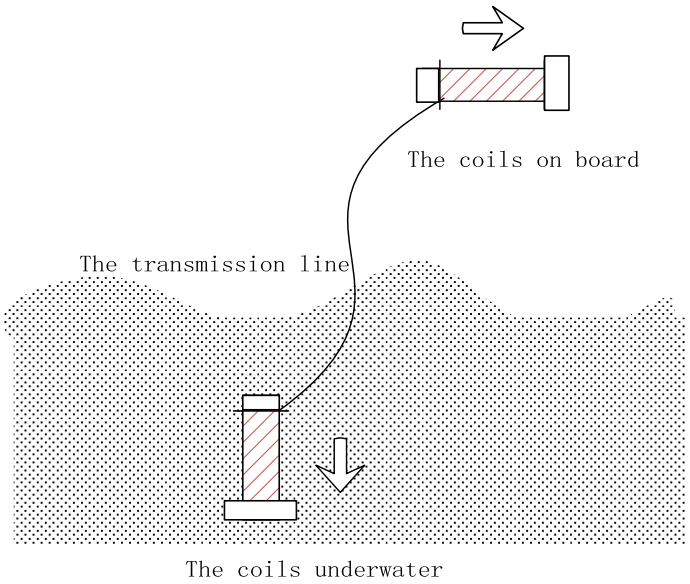


Fig. 1. The operation model of expendable measuring instruments

The transmission line—thin double-shield cable's diameter is only 0.1 mm. It was twisted on the spools of overwater and underwater device. With the process of data measuring, the line was released off the spools (Fig. 1).

The accuracy and efficiency of this underwater expendable measuring instrument transmission, however, is limited since several problems.

First, different from the state in the air, it is surrounded with the seawater then the distributed parameters, as well as the property of this information channel, will be changed since there is coupling between the line and seawater. Second, in the working procedure of this measuring instrument, key parameters of the channel are time-varying for both the length of the cable and the winding coils was changed. This is also an important influential factor in the accurate and high-speed transmission of the signal. Third, for such a long distance, if it exceeds 1000 m transmission line in a higher frequency, the signal integrity will be damaged bringing a huge difficulty for demodulation. In a word, this transmission line presents an extremely harsh environment for the communication signals. The three critical channel parameters, namely noise, impedance, and attenuation are found to be highly unpredictable which can make the signal unrecognized.

In the literature, many papers on the modeling of distribution parameters can be found, but these analyses are limited to a particular field and are not applicable to the

situation in this paper. Only, a few papers exists that cover the instance and capacitance calculation of marine communication channel. Zheng et al. [1] engaged in the modeling of such an undersea time-varying channel. More often, the distributed parameters calculation employed the method combining experimental results with other similar analysis, and it is still in the exploration stage.

The aim of this paper is to propose a new method for deriving inductance and capacitance of both undersea cable and coils for the expandable sea environment measuring equipment, and compare the theoretical and experimental results. The method is based on an analytical approach using the classical model of inductance and capacitance. In what follows, the channel was modeled, respectively, in cable part and coil part.

The paper is outlined as follows. After this introduction, the model of undersea channel and its extension will be revised briefly. The next chapter will demonstrate the analysis of cable capacitance undersea in detail. Section 4 explains the capacitance and inductance calculation of the "coils" of the undersea channel based on some assumption regarding the practical structure of the coils. The results of the approach described in Sects. 3 and 4 are verified by measurement of the capacitance and the inductance of four different coils in Sect. 5. Finally, a conclusion and topics of future research are given in Sect. 6.

2 Models of the Channel

The electromagnetic theory states that to achieve efficient point-to-point transmission of power and information, the source energy must be guided. When the lines are used to transmit high-frequency communication signals, they can be regarded as the transmission line, which guides the transverse electromagnetic waves along with them.

In this expendable measuring instrument, the cables are made up of the copper conductors with insulation material which relative dielectric constant is 4.5.

In the working procedure of expandable measuring instruments, the transmission channel can be approximated as a closed form of the "two-wire transmission line". According to the transmission line theory, with the elongated of the transmission line, the two-wire transmission cable must be a pair of parallel conducting wires separated by a uniform of distance, therefore, can be modeled based on a typical parallel two-wire modal. According to the [2], the coils both on board and undersea can be approximated as inductors. The inductor winding has distributed parasitic parameters, which can be modeled in a simple way by a lumped parameter equivalent circuit as shown in Fig. 2. It contains an inductance L, a series of AC resistance coils R_{ac}, and an overall stray capacitance C_S. Unlike the analysis in [2], in this paper, the analysis is performed for cylindrical inductors made of a uniformly wound double-parallel wires as described above but single wire. But the model of the inductors in [2] is still applied for the "coils" of expandable measuring instruments.

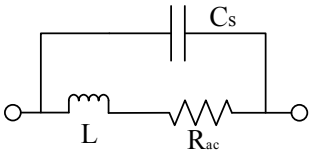


Fig. 2. The model of "coils"

In sum, the model of the communication channel can be divided into two parts: the cable spread out in water, and two "coils" in air and water. As the communication channel is surrounded with the seawater which is a good conductor coupled to the line, its main effect is imposing the stray capacitance on the all communication channel. So, in this case, the shunt capacitance should be given by the considerations from the turn-to-turn capacitance the layer-to-layer capacitance and the turn-to-water capacitance.

Based on the above consideration, the paired channel is regarded as a distributed parameters network, where the voltages and currents can vary in magnitude and phase over its length. Hence, it can be described by circuit parameters that are distributed over its length.

The communication channel modal in Fig. 3, for the two long-thin parallel wires part, R_o, L_o and C_o was defined as the resistance (in Ω/m), inductance (in H/m) and capacitance (in F/m) per unit length. For the "coils" part, as it is wounded by the two parallel wires, similarly, R_1, L_1 and C_1 defines the resistance(in Ω/m), inductance (in H/m) and capacitance (in F/m) per unit length, and C_{2o} is the capacitance between the cable and the seawater per unit length(in F/m), at last C_{12} is the distributed capacitance of one single wire itself per unit (in F/m). Considering the "coils" on aboard, similarly analyzing, the equivalent lumped-element circuit of the whole communication channel is described in Fig. 3.

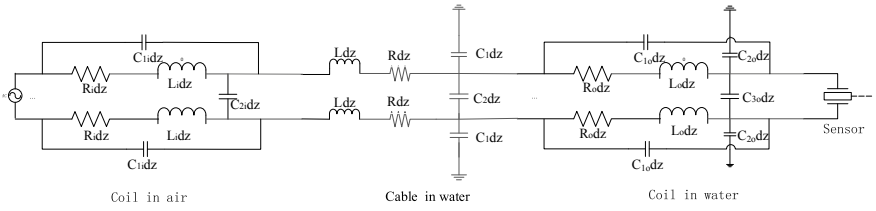


Fig. 3. The communication channel modal of expendable measuring instruments

3 Determination of Primary Line Parameters

In order to determine the performance of the channel, the primary parameters R, L, C both the line have to be determined first.

For the two long underwater parallel wires, in Fig. 3, this paper mainly describes the theoretical analysis of the C_o (capacitance between the two long parallel wires per unit distance), L_0 (inductance of the cable), and the C_{01} (capacitance between single line and seawater) which have a strong influence for the channel property.

3.1 Inductance (L0)

The inductance of the two-wire transmission line includes the self-inductance for each conductor and the mutual inductance between them. From total inductance, it can be obtained as

$$L_0 = 2L_s + L_m = 2 \times \frac{\mu_c}{8\pi} + \frac{\mu_c}{\pi} \ln\left(\frac{2R - r}{r}\right)(H/m) \tag{1}$$

where R is radius of one the conductors that include the insulating coatings, while r is the radius exclude the insulating coatings, L_s is self-inductance for one conductor defined by (2), and the Lm is the mutual inductance between a pair of parallel conductors defined by (3).

$$L_s = \frac{\mu_c}{8\pi}(H/m) \tag{2}$$

$$L_m = \frac{\mu_c}{\pi} \ln\left(\frac{2R - r}{r}\right)(H/m) \tag{3}$$

3.2 Capacitance Between the Two Conductors (C_0)

C_0 can be regarded as the capacitance of two-wire transmission line, which is given by [2] as

$$C1 = \frac{\pi\varepsilon}{\ln\left[\frac{r-(R-d)}{d+(R-r)}\right]} \tag{4}$$

Where ε is the relative permittivity of the dielectric material in between the conductors, and $d = \sqrt{R^2 - r^2}$.

3.3 Capacitance Between One Conductor and the Seawater (C_{01})

The determination for the C_{01} is more complicated since the cable and seawater are in the non-coaxial position and having an abstract shape. This kind of capacitance can be solved by the differential and integral calculus method multiplied with the classical method.

The cross-sectional geometric view between conductors and seawater is shown in Fig. 4. In a macrocosmic viewpoint, this geometric model can be approximated as two adjacent conductors for the elementary surfaces of these conductors dS which elementary capacitance dc can be given by $dc = \frac{\varepsilon dS}{x}$ where ε is the permittivity of medium and x is the length of a line of the electric field surface assuming it get out from a conductor fully surrounded by the seawater go to the seawater and no line can go to infinity. In particular, for this case, the surface between the conductor surface and the

seawater surface is cylindrical, shown in Fig. 5, and the elementary between related to the cylindrical coating shell is given by

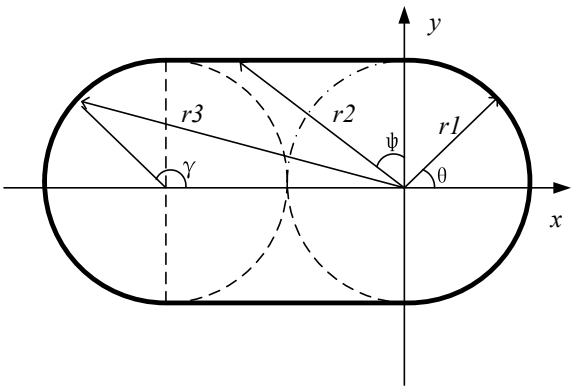


Fig. 4. The model between the sea and wire

$$\mathrm{dC} = \frac{\varepsilon r d\theta}{dr} \tag{5}$$

From the geometric model between the conductor and seawater, shown in Fig. 4, the radius r is not constant, but can be a function of the location of angle θ. Therefore, some coordinate system can be select. In this case, the value of each radius can be divided into three parts to describe by the Cartesian coordinate system, taking the center of any conductor as the origin, and then the relation between the angle θ and the radius r can be described by

$$\mathrm{r} = \begin{cases} r & -\frac{1}{2}\pi < \theta < \frac{1}{2}\pi \\ \frac{r}{\cos(\varphi)} & 0 < \varphi < arctan\left(\frac{b}{r}\right) \\ r\sqrt{1+\left(\frac{b}{r}\right)^2 - 2\times\frac{b}{r}\cos(\gamma)} & \frac{\pi}{2} < \gamma < \pi \end{cases} \tag{6}$$

In which b = 2R. Then, integrate this equation for r ranging from the radius of the conductor without the coating r to the out radius of the wire, including coating R. So, the total capacitance can be obtained as

$$\begin{aligned} C &= \varepsilon \int_r^R \frac{r}{dr}\left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} d\theta + 2\int_0^{\arctan\left(\frac{b}{r}\right)} \frac{1}{\cos\varphi} d\varphi + 2\int_{-\frac{\pi}{2}}^{\pi} \sqrt{9.35 - 5.78\cos\gamma}\, d\gamma\right) \\ &= \frac{\varepsilon}{\ln\left(\frac{R}{r}\right)}\left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} d\theta + 2\int_0^{\arctan\left(\frac{b}{r}\right)} \frac{1}{\cos\varphi} d\varphi + 2\int_{\frac{\pi}{2}}^{\pi} \sqrt{9.35 - 5.78\cos\gamma}\, d\gamma\right) \end{aligned} \tag{7}$$

It should be noted that when integrating this expression for φ and γ, they must transmitted the function of θ:

$$\begin{array}{ll} \gamma = \theta - \arcsin\left(\frac{2R}{r} sin\theta\right) & \arctan\left(\frac{r}{2R}\right) < \theta \leq \pi \\ \varphi = \frac{\pi}{2} + \theta & \frac{\pi}{2} < \theta \leq \arctan\left(\frac{r}{2R}\right) \end{array} \tag{8}$$

Since the capacitance calculation between the conductor and seawater essentially is a question-related capacitance calculation about the abstract shape and eccentric non-coaxial wires, so Eq. (5) can also be used to calculate the capacitance of any shape cross-session coaxial wires.

4 Determination of Primary Coils Parameters

For the "coils" wound by a pair of long-thin parallel conducting wires in this paper, the stored energy for a standard winding mainly concentrates between two adjacent layers. In comparison to that, the energy between two turns of the same layer is very small. Thus, the turn-to-turn capacitance (C_{12}) in the same layer is neglected in this paper. So, this section mainly analyses the layer-to-layer capacitance (C_1) and the layer-to-seawater capacitance (C_{11}). The calculation of layer-to-layer capacitance (C_1) is based on the static layer-to-layer capacitance according to the coils actual physical structure. And the calculation of layer-to-water capacitance (C_{11}) is based on the electrical energy stored between the seawater and the layer. At last the series inductance of this coil is also integrated based on the mutual inductance which is calculated using Neumann's formula and alternate methods.

4.1 Inductance (L_1)

According to the [3], for the inductance calculation of filamentary current paths in which the wire radius r is small compared with the loop radius R, the field is nearly the same as though current were concentrated along the center of the wire. Thus, the external inductance of this filamentary loop can be approximated by the mutual inductance between the two filaments in Fig. 6.

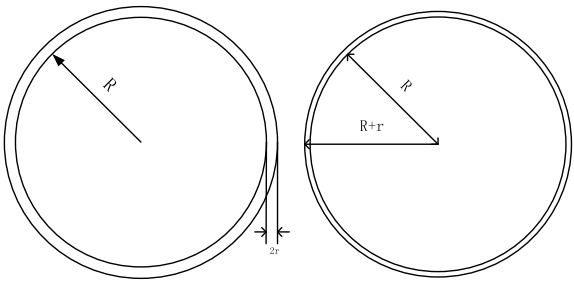


Fig. 5. (1) Round loop of wire (2) The approximated modal for inductance calculation

So, the inductance between two concentric circles of radius R and R-r using the Neumann's formula can be expressed as follows:

$$L_s = \mu(2R - r)[\left(1 - \frac{k^2}{2}\right)K(k) - E(k)] \tag{9}$$

Where $k^2 = \frac{4R(R-r)}{(2r-R)^2}$, R is the radius of the loop, and r is the radius of the wires. And if the r/R is very small, k is nearly unity and K and E may be approximated by

$$\begin{array}{c} K(k) \cong \ln\left(\frac{4}{1-k^2}\right) \\ E(k) \cong 1 \end{array} \tag{10}$$

So L_S can be calculated by

$$L_0 = r\mu[\ln\left(\frac{8R}{r}\right) - 2] \tag{11}$$

For a circular coil of N turns formed into a circular cross section, when the radius of wires is small compared with the coils radius, the calculation of self-inductance may modify the formula (9). First, magnetic field must be computed on the basis of a current NI but the current I, second to compute the total induced voltage about the coil, N integrations must be made about the loop. Thus, Eq. (9) is modified by a factor N^2, so the external inductance for this coil can be calculated by

$$L_1 = N^2 R\mu\left[\ln\left(\frac{8R}{r}\right) - 2\right] \tag{12}$$

4.2 The layer-to-layer capacitance

As shown in Fig. 7, the winding arrangements of "coils" can be divided into two types: Ortho-cyclic winding and Orthogonal winding which can be seen as the winding type of the "coils" in this paper since the "coils" were turned around by a pair parallel long thin wire. Furthermore in this section, the layer-to-layer capacitance is calculated by assuming that the layers of the winding are equipotential surface, i.e., that all turns of one layer are short-circuited thus the lines of the electric field must be orthogonal to the conductor surface and using the geometrical symmetry of the coils, edge effects and proximity effects are also neglected. Thus, two adjacent layers of the "coils" could be approximated by a parallel-plate capacitor.

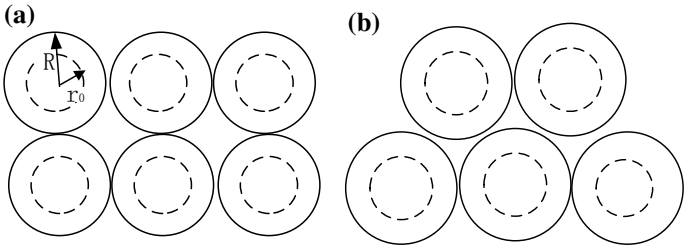


Fig. 6. Two kind of windings of coils

In literature, several techniques which are based on the static layer-to-layer capacitance have been introduced to calculating the stray capacitance of transformers that essentially is "coils". Basically, there are two approaches in capacitance calculation: the first one is based on energy method, the second one is based on the physical structure and analytical approach, for example, in this paper [4–6], they are all based on the second one, but in their analysis procedures although many processes are simplified are still too complex on microscopic lever considering the air gap between two insulating coatings and on macroscopic level considering the number of layers. In this paper, the approach is also based on the physical structure and analytical method, but because the thickness of the air gap in the practical "coils" is so much lower that can be negligible in the capacitance computing. And according to the "coils" practical structure, this section also presented the double-layer capacitance calculation model no matter how many layer is, shown in Fig. 8 which are more simplified and reduce the calculation workload.

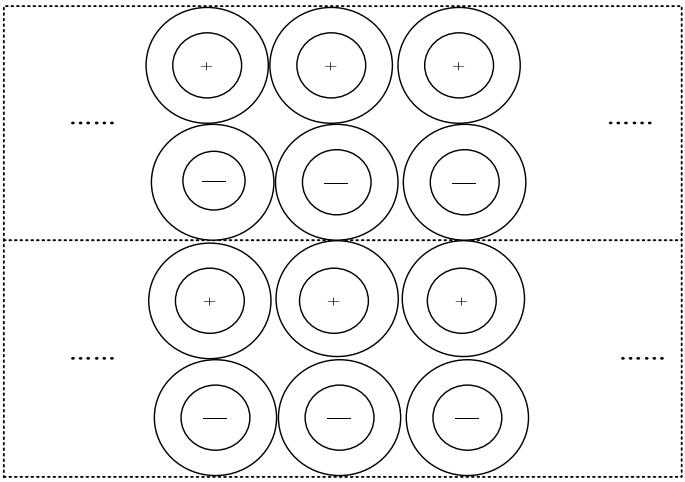


Fig. 7. Cross section view of coils

According to the calculation model shown in the Fig. 8, for orthogonal windings, the turns of successive layers are orthogonally on the top of each other which basic cell can show as Fig. 9. Approximating the electric lines as straight lines in the basic cell, it deviates toward the adjacent layer and runs as a straight line to upper and underlayer, respectively. And within one basic cell, all approximated electric flux lines from one turn to the underlying one are contained which is used for all angles φ from 0 to π. Again, in Fig. 10, the resulting flux lines of a finite element analysis are shown. The FEA-flux lines correspond quiet well with the approximated flux line especially in the regions where the stored electric energy and therefore parasitic capacitance is high (ϕ is around $\frac{\pi}{2}$).

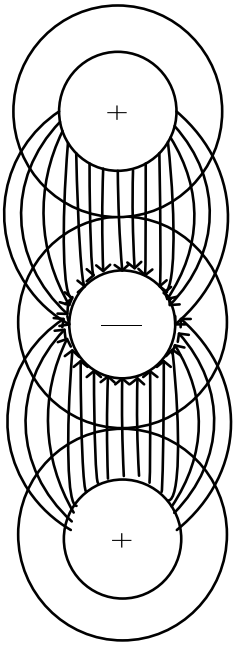


Fig. 8. Basic cell for capacitance calculating

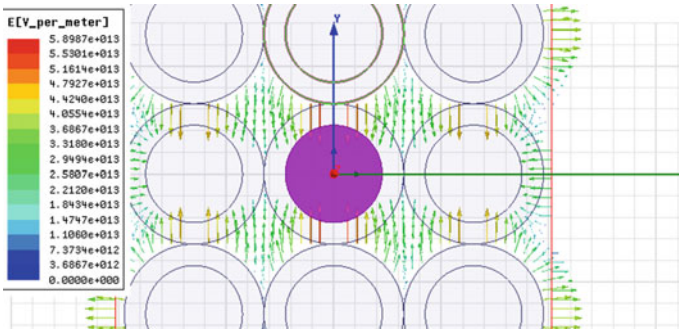


Fig. 9. Electric field line of basic cell based on FEA

With the considered path of flux lines, a given permittivity, as well as thickness insulation, an average turn surface of adjacent layers, the static layer-to-layer capacitance, could be calculated. The first one replaces the round wires by flat rectangles, which can be replaced by an equipotential if the electrical connection between the layers is disconnected and dc conditions are assumed. Finally, using the parallel-plate capacitor basic capacitance calculation formulas $dC = \frac{\varepsilon dS}{d}$, replacing the distanced by the thickness of insulation between two adjacent layers, considering the geometric

structure the practical basic cell contains such parallel capacitors, layer-to-layer capacitance in the basic cell could be calculated as

$$\mathrm{C}_1 = \varepsilon \frac{4\pi^2 Rr}{d} \tag{13}$$

Multiplying this capacitance by the number of turns results in the static capacitance of the layer-to-layer capacitance:

$$\mathrm{C} = \mathrm{N} \times 2 \times \mathrm{C}_1 = \varepsilon \frac{8\pi^2 Rr}{d} \mathrm{N} \tag{14}$$

4.3 The layer-to-seawater capacitance

According to the applied winding structure, the effective capacitance changes because the static capacitance is not charged uniformly but the charges distribution depend on the voltage distribution which is depending on the winding method. Assuming that the "coils" winding method in this paper is standard.

If it is assumed that all turns between the layer and seawater comprise the same magnetic flux the voltage distribution along the layer as shown in Fig. 5 is approximately linear, so the voltage distribution between the layer and seawater can be described as $\mathrm{U(k)} = \frac{U}{N}$. Since the average voltage between the seawater and the layer is $U_1 = \frac{U}{2}$, the energy stored between the layer and seawater can be calculated as $W = \frac{1}{2} C U_1^2 = \frac{1}{8} C U^2$, resulting in the layer-to-seawater capacitance is $C = \frac{8W_e}{U^2}$.

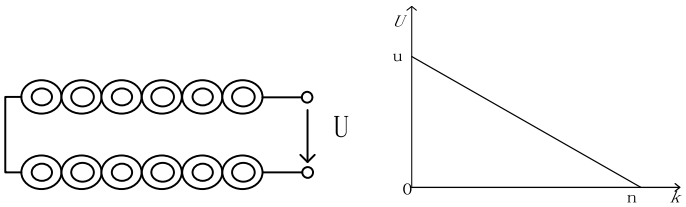


Fig. 10. The winding of single wire and its voltage distribution

On the other hand, for a linear homogeneous medium region, the system energy can also described as $W_e = \frac{1}{2} \iiint \vec{E} \cdot \vec{D} dV$, with the $\vec{D} = \varepsilon \cdot \vec{E}$, the energy stored in the element of turns dk and seawater can be calculated as:

$$\mathrm{dW}_{\varepsilon} = \frac{\varepsilon_0 \varepsilon_r}{2d^2} U(k)^2 lSdk \tag{15}$$

Integrating this equation for k ranging from the 0 to the N, the energy of whole region between the seawater and "coil" can be expressed as

$$W_{\varepsilon} = \int_0^N \frac{\varepsilon_0 \varepsilon_r}{2d^2} U(k)^2 lSdk = \frac{\varepsilon_0 \varepsilon_r NLSU^2}{6d^2} \tag{16}$$

Combined with the expression $C = \frac{W_e}{U^2}$, the layer-to-seawater capacitance is

$$C = \frac{4\varepsilon_0 \varepsilon_r lNS}{3d^2} \tag{17}$$

5 Experimental Verification

To verify the correctness of the calculations, the parameters of both the cable and "coils" were measured with a U2818 LCR meter at different frequency from 0 to 20 MHz. The coils were carefully wound on the plastic tube from enameled copper wire, turn by turn.

The results given by the proposed method have been compared against those measured for several "coils" and cables.

Table 1 shows the computed and measured values of a pair of parallel conducting wires whose length L = 1 m, outer radius include insulation R = 0.65 mm (for one conductor), inner radius without coating r = 0.45 mm, and the center distance of the two conductor d = 1.3 mm, and the medium permittivity ε = 4.5, thickness s = 0.4 mm. There were no significant deviations between the measured and calculated values and the deviation majority of the values usually does not exceed 1%.

Table 1. Measured and computed values of the parameters for parallel wires

Parameters	Measured value	Calculated value	Relative deviation
Line-to-line capacitance	93.5pF	95.20pF	0.01
Line-to-seawater capacitance	685pF	686.16pF	0.001

Table 2 shows the computed and measured values of "coils" with the pair of parallel conducting wires described above wound on the plastic tube using the standard method, and its diameter D is 18 mm and the number of turns N is 150. And similarly, there were no significant deviations which usually does not exceed.

Table 2. Measured and computed values of the parameters for underwater coils

Parameters	Measured value	Calculated value	Relative deviation
Layer-to-layer capacitance	2120pF	2386pF	0.11
Layer-to-seawater capacitance	7450pF	7782pF	0.04
Inductance	470uH	513.42uH	0.09

Furthermore, to verify the accuracy of the parameter calculation for "coils" other three different diameter and different number turns samples are also measured and compared with the computed values, all in all, the relative error is within 15% in the degree that we can tolerant approving that the theoretical analysis is usable.

6 Conclusion

In this paper, the communication channel for expandable measuring instruments is modeled combining the transmission line theory with the effect of seawater. A serials method for predicting the crucial parameters of the whole channel have been proposed herein. Experimental measurements also have been introduced to verify the computed values. The results are very promising and show a good compromise between the computed value and measured value. New contributions of this work are as follows

- For calculating the layer-to-layer capacitance, the double-layer capacitance computing model which can reduce the entire calculation process has been introduced and verified. This method can also be used for the capacitance analysis of other coils for different applications.
- For calculating the capacitance between the conductor and seawater, a new method has been presented and verified. This method is reasonable for any shape cross-section coaxial wires capacitance calculation.
- The method for calculating the inductance of coils has been shown and verified, and provide the guidance for other coil inductance calculation.

In summary, this paper present a new method for predicting the critical parameters of the communication channel of vessel carried undersea expendable measuring instruments and also verifying with experimental measurements. And, there were small deviations in the comparing results, approving that the analysis method is feasible, effective, and suitable for the channel analysis. Finally, the work in this paper lays a solid foundation for further channel performance research, as well as the establishment of time-varying channel modal and also provides references for other undersea channel research.

Acknowledgements. This paper's work is supported by the National Key Research and Development Program of China (Grant No. 2016YFC1400501).

References

1. Zheng, Y., Zhao, X., et al.: Analysis of time-varying channel effects on transmission performance of abandoned measuring instrument. Acta. Phys. Sin. **64**(4), 1–7 (2014) (in Chinese)
2. Xie, C., Rao, K.: Electromagnetic Field and Waves. Higher Education Press, Beijing (2006) (in Chinese)
3. Duerdoth, W.T.: Equivalent capacitance of transformer windings. Wirel. Eng. **23**(6), 99–119 (1946)

4. Koch, J.: Berechnung der kapazität von spulen insbesondere in schalenkernen. Valvo Berichte. **24**(6), 671–676 (1922)
5. Massarini, A., Kazimierczuk, M.K.: Self-capacitance of inductors. IEEE Trans. Power Deliv. **12**(4), 671–676 (1997)
6. Gandi, G., Kazimierczuk, M.K., et al.: Stray capacitance single-layer spool air-core inductors. IEEE Trans. Ind. Appl. **35**(6), 1162–1168 (1999)

Analysis of Timing Performance of Zadoff–Chu Sequence in WFRFT Domain under Frequency Offset Effect

Zhongyao Chen(✉), Xuejun Sha, and Xuejiao Liu

School of Electrical and Information Engineering, Harbin Institute of Technology, Harbin, China
chen_zhongyao@163.com

Abstract. Zadoff–Chu (ZC) sequence is widely used as synchronization sequences in modern communication systems for its perfect autocorrelation properties. We first investigate the correlation properties of ZC sequence after weighted fractional Fourier transform (WFRFT) operation. We then analyze how WFRFT operation affect the timing synchronization performance under the presence of carrier frequency offset (CFO) through timing spectrum. Demonstrated through simulation, the ZC sequence in WFRFT achieve better timing performance while CFO exits.

Keywords: Zadoff–Chu sequence · Correlation function · Synchronization · CFO · WFRFT

1 Introduction

Synchronization is a very important issue in digital communications, which is the very first step in establishing a communication link to resolve the timing and frequency uncertainties between the transceivers. It is typically achieved by using training sequences in the frame header in modern wireless communication systems. Compared to the schemes with cyclic prefix aided, the training sequence can always provide higher accuracy in timing performance.

Zadoff–Chu (ZC) sequences [1], a class of polyphase sequences, is well known for its ideal autocorrelation property, which makes it very suitable for time synchronization in modern wireless communication systems. In LTE systems, ZC sequences have been used as primary synchronization signals (PSS), replacing the conventional pseudo-random noise sequences due to their perfect autocorrelation properties. Although synchronization schemes using ZC sequences have been extensively studied, most of them are based on the assumption of no frequency offset between the transmitter and the receiver. The ideal periodic autocorrelation property of ZC sequences is also under this assumption. Unfortunately,

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 171–179, 2019.
https://doi.org/10.1007/978-981-13-6264-4_20

when carrier frequency offset (CFO) exists, the effect of CFO on autocorrelation property of ZC sequences cannot be ignored. The perfect time autocorrelation is no longer true in practical applications when a frequency synchronization error is present between the transmitter and the receiver. A nonzero CFO actually causes energy leakage of the correlation peak and even shift the peak to a wrong timing position under large CFO. Recently, the effect of CFO on the autocorrelation of the ZC sequences was investigated. Reference [2] developed an analytical framework, particularly timing spectrum, which fully characterizes a ZC sequences timing properties and its fundamental limitations as a time synchronization sequence in the presence of CFO. Taking the CFO effect into account, the timing performance of ZC sequences as synchronization signal is not that ideal as its periodic autocorrelation without CFO.

This current paper focuses on investigating the timing synchronization performance of ZC sequences in WFRFT domain under the presence of CFO. Our main contributions are as follows. First, a systematic analysis of the correlation between the received preamble and local template sequence is performed. Then, we use the timing spectrum to reveal the critical frequency offset of the proposed sequence used as a synchronization sequence. Third, simulation results obtained to verify that the ZC sequence in WFRFT domain significantly outperforms the conventional ZC sequence in time domain under various CFOs.

2 Theoretical Basis

2.1 Zadoff–Chu Sequence

ZC sequence belongs to the class of constant amplitude zero autocorrelation sequences [3]. Such sequences are favored for synchronization purposes because they have a constant amplitude and exhibit the useful property that cyclically shifted versions of a ZC are orthogonal to one another. Furthermore, the discrete Fourier transform of a ZC sequence is another ZC sequence [4]. A ZC sequence of length N is defined as

$$s_\mu(n) = e^{-j\frac{\pi\mu n(n+1)}{N}}, \quad n = 0, 1, \ldots, N-1 \tag{1}$$

where μ is the root index chosen to be relatively prime with respect to length N. The number of available ZC sequences of length N equals the number of integers that are relatively prime to N. When N is an odd prime, the maximum number of distinct ZC sequences of length N exists. Hence typically N is chosen to be an odd prime, which we assume in this paper.

In some conditions, however, the required length M of ZC sequence is not a prime integer, and hence prime-length ZC sequences cannot be directly used. Instead, it is derived from a truncated ZC sequence of length N, where N is the smallest prime number $\geqslant M$.

ZC sequences possess good correlation properties which are essential in timing synchronization applications between a UE and a base station. The periodic autocorrelation function is defined as

$$R_{ss}(\tau) \triangleq \frac{1}{N}\sum_{n=0}^{N-1} s_{\mu}(n)\, s_{\mu}^{*}((n-\tau)_N) = \begin{cases} 1, & \tau = 0 \\ 0, & otherwise \end{cases} \tag{2}$$

where $(\cdot)_N$ denotes the modulus-L operation, $*$ is complex conjugation. Clearly, the ZC sequence has an ideal cyclic autocorrelation property. In addition, the periodic cross-correlation function of two ZC sequences s_{μ_1} and s_{μ_2} of length N is defined as

$$R_{s_{\mu_1}s_{\mu_2}}(\tau) \triangleq \frac{1}{N}\sum_{n=0}^{N-1} s_{\mu_1}(n)\, s_{\mu_2}^{*}((n-\tau)_N) = \pm\frac{1}{\sqrt{N}}, \quad \mu_1 \neq \mu_2 \tag{3}$$

for any integer τ. The correlation properties of ZC sequences remain invariant under cyclic shifts or addition of constant phase shift in exponentials.

In Ref. [2], it was observed that there exists a duality between ZC sequences in time domain and frequency domain. Reference [4] has given the proof that the energy-normalized N-point discrete Fourier transform (DFT) of a time domian ZC sequence is defined as

$$\begin{aligned} S_{\mu}(k) &= \frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} s_{\mu}(n) e^{-j\frac{2\pi nk}{N}} \\ &= S_{\mu}(0)\cdot s_{\mu}^{*}(\mu' k), \quad k = 0, \ldots, N-1 \end{aligned} \tag{4}$$

where μ' is defined such that $\mu'\mu \bmod N = 1$. Note that $S_{\mu}(k)$, the DFT of a time domain ZC sequence is still a ZC sequences with ideal correlation properties. Hence, under actual applications, a ZC sequence can be directly generated in the time or frequency domain without the need for a DFT/IDFT operation.

2.2 Defination of Weighted Fractional Fourier Transform

Weighted fractional Fourier transform is a new mathematics tool for time–frequency analysis, which has a variety of definitions. The N-point α-th order 4-WFRFT of any discrete time signal $x(n)$ is define as [5] a linear summation of discrete time signal $x(n)$ and its discrete Fourier transform $X(n)$. 4-WFRFT is expressed by

$$F_{4W}^{\alpha}[x(n)] = \omega_0(\alpha)x(n) + \omega_1(\alpha)X(n) + \omega_2(\alpha)x(-n) + \omega_3(\alpha)X(-n) \tag{5}$$

where F_{4W}^{α} denotes the WFRFT operator with 4 weighted functions, and $X(n)$ is the DFT of $x(n)$. $x(-n)$ and $X(-n)$ are the inverse order sequences of $x(n)$ and $X(n)$. DFT operation is implemented as

$$X(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi nk}{N}}, \quad k = 0, \ldots, N-1 \tag{6}$$

The weighted parameters are denoted by

$$\omega_l(\alpha) = cos\left[\frac{(\alpha - l)\pi}{4}\right] cos\left[\frac{2(\alpha - l)\pi}{4}\right] exp\left[\pm j\frac{3(\alpha - l)\pi}{4}\right] \quad l = 0, 1, 2, 3 \tag{7}$$

where the sole crucial parameter α is a real number. As $\omega_l(\alpha) = \omega_l(\alpha + 4)$, the parameter α is chosen in $[0, 4)$.

According to the analysis of the relationship between weighted parameters in (7), four functions of the integer order WFRFT are given

$$\begin{cases} F_{4W}^{0}\left[x(n)\right] = x(n) \\ F_{4W}^{1}\left[x(n)\right] = X(n) \\ F_{4W}^{2}\left[x(n)\right] = x(-n) \\ F_{4W}^{3}\left[x(n)\right] = X(-n) \end{cases} \tag{8}$$

The complexity of α-th order WFRFT equals to that of fast Fourier transform.

3 Properties of ZC Sequence in WFRFT Domain

The weighted fractional Fourier transform of a ZC sequence s_μ of length N is given by

$$F_{4W}^{\alpha}\left[s_\mu(n)\right] = \omega_0(\alpha)s_\mu(n) + \omega_1(\alpha)S_\mu(n) + \omega_2(\alpha)s_\mu(-n) + \omega_3(\alpha)S_\mu(-n) \tag{9}$$

Due to properties of ZC sequence described in the previous section, $S_\mu(n)$ can be replaced by Eq. (4)

$$\begin{aligned} s_\mu^{\alpha}(n) &= \omega_0(\alpha)s_\mu(n) + \omega_1(\alpha)S_\mu(0)s_\mu^*(u'n) \\ &\quad + \omega_2(\alpha)s_\mu(-n) + \omega_3(\alpha)S_\mu(0)s_\mu^*(-u'n) \\ &= \omega_0(\alpha)s_\mu(n) + \omega_1(\alpha)s_\mu^*(u'n)e^{j\phi} \\ &\quad + \omega_2(\alpha)s_\mu(-n) + \omega_3(\alpha)s_\mu^*(-u'n)e^{j\phi} \end{aligned} \tag{10}$$

where s_μ^{α} is a new sequence generated by the WFRFT of a ZC sequence with root index μ, determined by two parameters: μ (root index) and α (order of WFRFT).

The periodic autocorrelation function is expressed as

$$R_\alpha(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} s_\mu^{\alpha}(n)\, s_\mu^{\alpha *}\left((n - \tau)_N\right) \tag{11}$$

When $\tau = 0$, which means no cyclic shift, clearly, $R_\alpha(0) = 1$ because WFRFT does not change the energy of signal. However, when $\tau \neq 0$, the autocorrelation function gets more complex. Replace s_μ^α with Eq. (10) and split the weighted component of s_μ^α,

$$\begin{aligned} R_\alpha(\tau) = \frac{1}{N}\sum_{n=0}^{N-1} & [\omega_0 s_\mu(n) + \omega_1 s_\mu^*(u'n)e^{j\phi} + \omega_2 s_\mu(-n) \\ & + \omega_3 s_\mu^*(-u'n)e^{j\phi}][\omega_0(\alpha)s_\mu(n-\tau) + \omega_1 s_\mu^*(u'(n-\tau))e^{j\phi} \\ & + \omega_2 s_\mu(-(n-\tau)) + \omega_3 s_\mu^*(-u'(n-\tau))e^{j\phi}] \\ & \leq \frac{2}{\sqrt{N}}(\omega_0\omega_1 + \omega_0\omega_3 + \omega_1\omega_2 + \omega_2\omega_3) \\ & + \omega_0\omega_2\delta(\tau-1) + \omega_0\omega_2\delta(\tau+1) \\ & + \omega_1\omega_3\delta(\tau-\mu) + \omega_1\omega_3\delta(\tau+\mu) \end{aligned} \tag{12}$$

where τ is an integer in $[-N/2, N/2]$, $\delta(\tau)$ denotes normalized impluse function. Obviously, When $\tau \neq \pm 1, \pm\mu$,

$$R_\alpha(\tau) \leq \frac{2}{\sqrt{N}}(\omega_0\omega_1 + \omega_0\omega_3 + \omega_1\omega_2 + \omega_2\omega_3), \quad \tau \neq \pm 1, \pm\mu \tag{13}$$

It is easy to show that the ZC sequence in WFRFT domain no longer remains the ideal autocorrelation porperty, instead, gets multi-valued porperty under $\tau \neq 0$ jointly determined by α and τ. Similarly, the periodic cross-correlation function of two ZC sequences s_{μ_1} and s_{μ_2} of length N is defined as

$$\begin{aligned} R_{\alpha,\mu_1,\mu_2}(\tau) = \frac{1}{N}\sum_{n=0}^{N-1} & [\omega_0 s_{\mu_1}(n) + \omega_1 s_{\mu_1}^*(u_1'n)e^{j\phi} + \omega_2 s_{\mu_1}(-n) \\ & + \omega_3 s_{\mu_1}^*(-u_1'n)e^{j\phi}][\omega_0(\alpha)s_{\mu_2}(n-\tau) + \omega_1 s_{\mu_2}^*(u_2'(n-\tau))e^{j\phi} \\ & + \omega_2 s_{\mu_2}(-(n-\tau)) + \omega_3 s_{\mu_2}^*(-u_2'(n-\tau))e^{j\phi}] \\ & \leq \frac{1}{\sqrt{N}}(\omega_0+\omega_1+\omega_2+\omega_3)(\omega_0+\omega_1+\omega_2+\omega_3) \end{aligned} \tag{14}$$

for any integer τ.

Here, we use detailed examples to help better understand and validate the analytical results developed in this section. We use two exemplary ZC sequences (one with = 140 and one 367, both of length $N = 839$ which is the same length used in LTE RACH) to show their distinctive correlation properties. Figure 1 shows the cyclic autocorrelation for ZC sequences $\mu = 140$ in $\alpha = 0.5$ WFRFT domain.

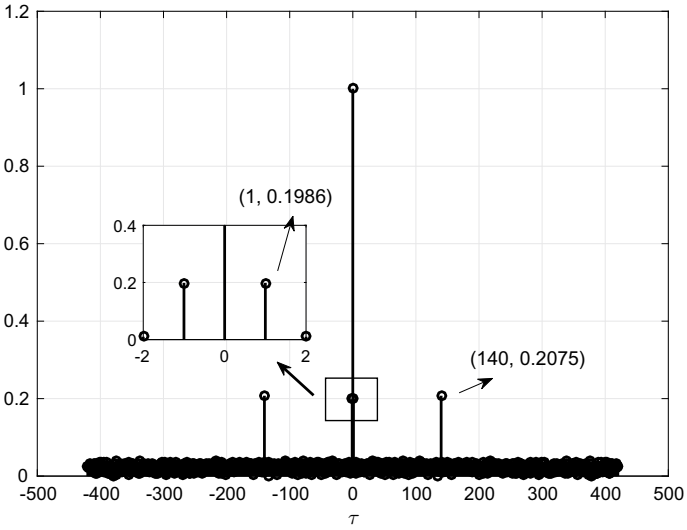


Fig. 1. Autocorrelation for ZC sequences $\mu = 140$ in $\alpha = 0.5$ WFRFT domain

From the figures shown above, the WFRFT operation does weaken the original ideal perfect autocorrelation, especially in the critial timing offset ± 1 and $\pm\mu$ as the Eq. (12) shown above, however, the autocorrelation of ZC sequence in WFRFT domain is still sharp enough for applications in timing synchronization.

In Ref. [2], a powerful analytical framework, called timing spectrum, was developed to characterize the timing properties of a ZC sequence under the effect of CFO. Figure 2 gives the autocorrelation of the ZC sequence $\mu = 140$ as a function of the normalized frequency offset ϵ at the specific shift offsets $\Delta\theta \in [-15, 15]$ associated with a timing observed window size of 15. The black solid curve in Fig. 2 represents the law of change of autocorrelation value at zero shift offset $\Delta\theta = 0$ with the normalized frequency offset changes. Other colored curves represent the law of change of autocorrelation value with the normalized frequency offset at nonzero shift offset $\Delta\theta \neq 0$. Clearly, at a small interval around the critical frequency offset, such as $\epsilon = \pm 1$, any small frequency offset incurs a large autocorrelation value at critical shift offset $\Delta\theta = \pm 6$. This value becomes even greater than that at $\Delta\theta = 0$ when frequency offset $|\epsilon| > 0.5$. Under such circumstance, although without the effect of noise, the detector will misjudge the start timing position of the sequence. The frequency offset that ZC sequence can bear has a theoretical limit $[-0.5, 0.5]$.

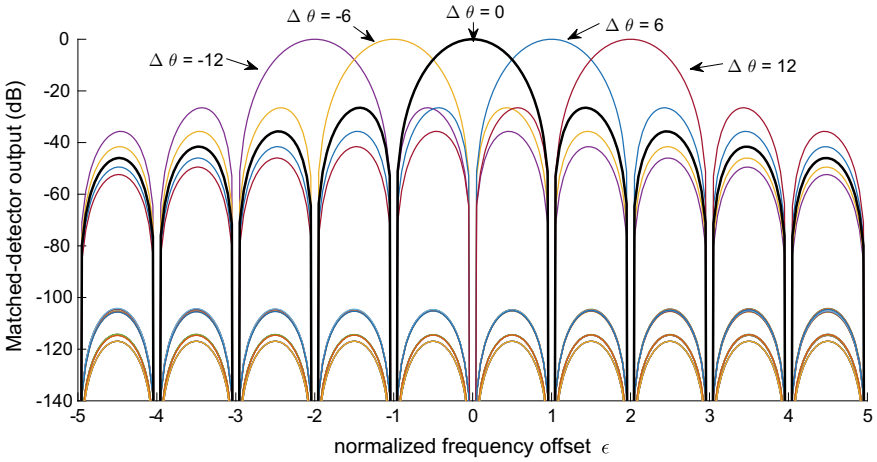


Fig. 2. Autocorrelation timing spectrum for ZC sequence $\mu = 140$ at shift offset $\Delta\theta \in [-15, 15]$ as a function of normalized frequency offset ϵ.

Figure 3 plots the autocorrelation timing spectrum of the same ZC sequence we discussed before but in $\alpha = 0.8$ WFRFT domain. We can see that, compared to timing spectrum of ZC sequence in time domain, the four timing spectral components at the critical frequency offset are weakened after WFRFT operations. Within a larger frequency offset range than 0.5, the timing spectral component at $\Delta\theta = 0$ covers other timing spectral component, has the maximum value of the autocorrelation. Meanwhile, at a specific frequency offset near 0, the

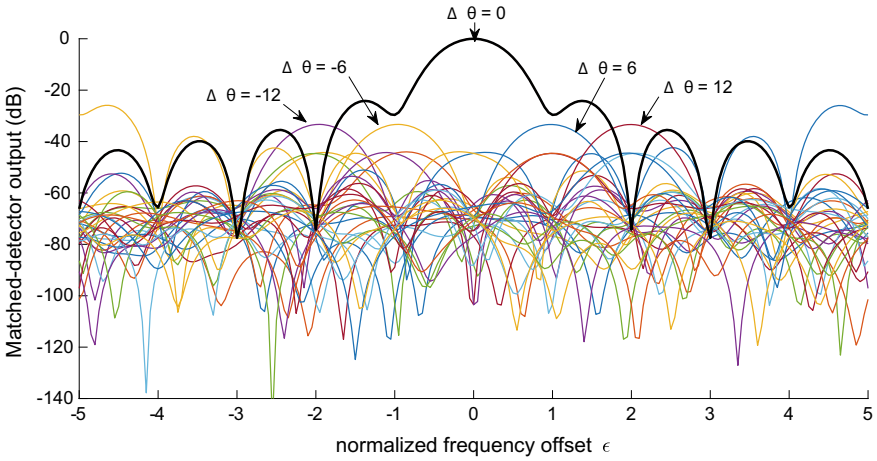


Fig. 3. Autocorrelation timing spectrum for ZC sequence $\mu = 140$ in $\alpha = 0.8$ WFRFT domain at shift offset $\Delta\theta \in [-15, 15]$ as a function of normalized frequency offset ϵ.

numerical distance in the vertical direction between the timing spectral components at $\Delta\theta = 0$ and $\Delta\theta \neq 0$. Better timing performance from ZC sequence in WFRFT domain can thus be expected.

4 Simulation Result

A numerical simulation was conducted to verify the timing performance of the new sequence from ZC sequence in WFRFT domain. The transmitted preamble has $N = 839$ samples, generated from ZC sequence $\mu = 140$ in $\alpha = 0.8$ WFRFT domain. The maximum likelihood estimation (MLE) is performed to estimate the signal timing, to determine the timing offset θ, which is based on the circular correlation of received signal $r(n)$ and the conjugated version of the original training sequence $x(n)$. In Fig. 4, we plot the curves of timing error probabilities at four different various CFO values $\epsilon = 0.3, \epsilon = 0.45, \epsilon = 0.6, \epsilon = 1.2$ under certain signal-to-noise ratio (SNR). It can be seen from Fig. 4, that compared with the ZC sequence in time domain, the ZC sequence in WFRFT domain can achieve a lower timing error probability when CFO exists. Moreover, the ZC sequence in WFRFT domain is functional in a larger frequency offset range, breaking the limits of frequency offset of ZC sequence in time domain.

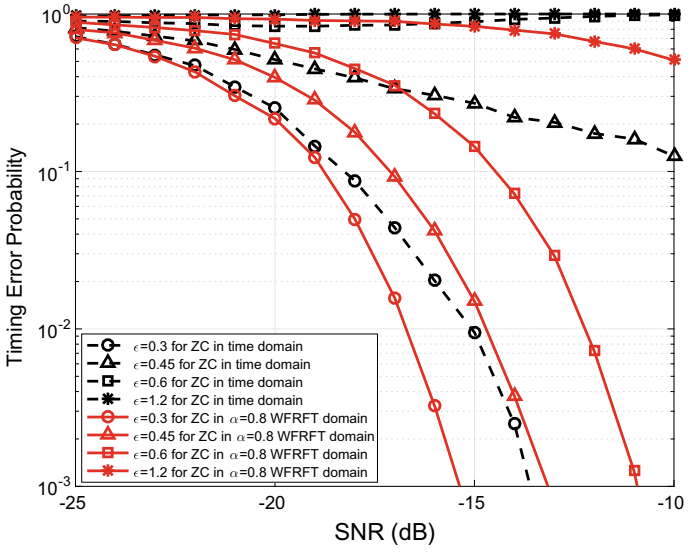


Fig. 4. Simulated timing error probabilities at various frequency offsets.

5 Conclusion

In this paper, we analyze the autocorrelation and cross-correlation properties of ZC sequence in WFRFT domain. Give an explanation on how the ZC sequence in WFRFT domain can achieve better performance through timing spectrum. In addition, we simulate the timing synchronization error probabilities at various CFO. Simulation results have demonstrated the timing synchronization performance of ZC sequence in WFRFT domain.

Acknowledgments. This work was supported by the National Nature Science Foundation of China under Grant 61671179.

References

1. Chu, D.: Polyphase codes with good periodic correlation properties (Corresp.). IEEE Trans. Inf. Theory **18**(4), 531–532 (1972)
2. Hua, M., Wang, M., Yang, K.W., et al.: Analysis of the frequency offset effect on Zadoff-Chu sequence timing performance. IEEE Trans. Commun. **62**(11), 4024–4039 (2014)
3. Popovic, B.M.: Generalized chirp-like polyphase sequences with optimum correlation properties. IEEE Trans. Inf. Theory **38**(4), 1406–1409 (1992)
4. Mansour, M.M.: A Hardware-Efficient Algorithm for Real-Time Computation of Zadoff-Chu Sequences. Kluwer Academic Publishers (2013)
5. Mei, L., Sha, X., Ran, Q., et al.: The research on the application of 4-WFRFT in communication system. Sci. China **53**(6) (2010)
6. Kang, J.W., Whang, Y., Ko, B.H., et al.: Generalized cross-correlation properties of Chu sequences. IEEE Trans. Inf. Theory **58**(1), 438–444 (2012)
7. Carni, E., Spalvieri, A.: Synchronous CDMA based on the cyclical translations of a CAZAC sequence. IEEE Trans. Wirel. Commun. **8**(3), 1144–1147 (2009)
8. Tao, J., Yang, L.: IImproved Zadoff-Chu sequence detection in the presence of unknown multipath and carrier frequency offset. IEEE Commun. Lett. **PP**(99), 1 (2018)

Modeling and Optimization for Sky Wave Propagation Loss Model Based on Enumerative Algorithm

Yuming Gu[1(✉)], Zhenyu Na[1], Xiao Liu[1], Huahan Zhang[1], and Wei Wang[2]

[1] School of Information Science and Technology, Dalian Maritime University, Dalian 116000, China
gym890@outlook.com, nazhenyu@dlmu.edu.cn, birdylx96@gmail.com, DMUmaple@163.com

[2] College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin 300000, China
weiwang@tjnu.edu.cn

Abstract. Sky wave propagation is one of the important methods to transmit signal and is widely used in ship broadcast communication. In order to analyze the multi-hop system and optimize the emission angle of sky-wave communication, this paper mainly models the transmission loss and optimizes the sky wave propagation path based on the enumerative algorithm. The simulation result shows the optimization scheme is useful for improving the signal-to-noise ratio (SNR) and quality of ship communication.

Keywords: Multi-hop system · Reflection loss · Enumerative algorithm · Optimization

1 Introduction

The long-distance communication is mainly achieved via the transmission of base station, satellite, and sky-wave communication. However, a center as a transfer device is both needed in base station communication and satellite communication, which is likely to be destroyed during the war and cause interruption of communication [1]. Therefore, in the past several years, the sky wave transmission has been widely used and studied for ship communication and military purpose because of its simplicity and stability [2]. And it is worthwhile for us to explore and optimize the propagation process [3].

For the process of sky wave transmission, the researches have given rise to many explorations to complete the transmission process [4]. Sky wave propagation loss is composed of transmission loss and natural attenuation [4]. The transmission loss is exploring the loss during the whole transmission [5]. Currently, the researches on transmission loss mainly focus on the following aspects: free space loss, ionosphere loss and extra system loss [6, 7].

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 180–187, 2019.
https://doi.org/10.1007/978-981-13-6264-4_21

In this paper, we not only explore the reflection loss which is another potential factor leading to propagation loss, but refine the entire propagation process based on the enumeration algorithm as well. First, we establish a Pierson–Moscowtz (PM) sea spectrum model, based on which, we tentatively investigate the reflection loss of sea surface. Second, according to the search algorithm, we determine the emission angle of high-frequency signals and complete the simulation of propagation. Finally, simulation results validate the effectiveness of enumeration algorithm.

2 System Model

The overall system proposed in this paper is composed of two parts: the transmission loss model and the optimization of emission angle (Fig. 1).

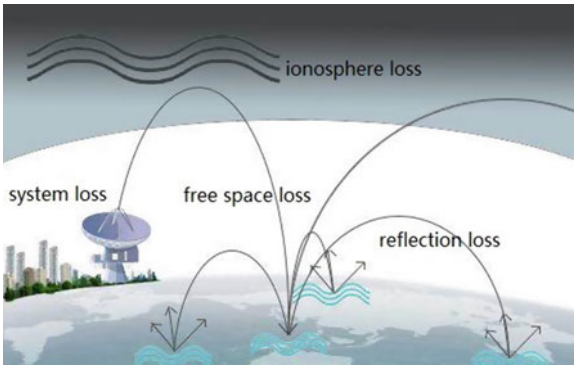


Fig. 1. Transmission loss model

Sky-wave field strength is usually calculated according to two methods: the corresponding parameters table and the empirical formulas are given by

$$E = 137.2 + 20lgf + p + G - L_b. \tag{1}$$

where f is the working frequency (Hz); p is power (kW) which is usually a given value; G is the gain factor; L_b is the transmission loss. In this paper, we do not discuss G, f and p because these parameters are given in constant values. Then, we have to explore L_b in detail. Therefore, we build the transmission loss model which is mainly composed of several losses and that is shown in formula (2).

$$L_b = L_o + L_a + L_g + L_p \tag{2}$$

where L_o is free space loss; L_a is ionospheric absorption loss; L_g is surface reflection loss; L_p is extra system loss which can often be neglected in this problem.

3 Proposed Scheme

3.1 Transmission Loss Model

The transmission loss is the main part which needs to be discussed. In this section, we consider a number of possible losses resulting from the transmission. There are several reasons causing the loss L_b:

$$\begin{cases} L_b = L_o + L_a + L_g + L_p \\ L_o = n \cdot \dfrac{2h}{\sin\theta} \cdot \rho \\ L_a = \sin\theta \cdot L_{a_0} \\ L_g = (n-1)L_{g_0} \end{cases} \tag{3}$$

where n is the hop times, θ is emitted angle, h is the elevation of D level (one of the ionospheric level, ρ is the loss of free space propagation of per unit distance, L_{a_0} is the ionospheric loss at an emission angle of 90°, L_{g_0} is sea surface refraction of every single loss.

Since the reflection loss is not considered by other researchers, we need to discuss it. First, we use the PM spectrum to simulate the ocean surface. The ocean surface can be seen as the overlying of several computational sine waves with different amplitudes, directions, frequencies and random phases. Thus, the formation of wave can be seen as a random process, which can be calculated by

$$\mathrm{S}(\omega) = \alpha \frac{g^2}{\omega^5} exp\left[-\beta\left(\frac{g}{U_{19.5}\omega}\right)^2\right] \tag{4}$$

Second, we filter the PM spectrum in frequency domain, and then do the inversion of Fourier transform to preserve the characteristic of surface's elevation so as to extract the characteristic of ocean surface. Assuming that a random position on the simulated sea surface can be expressed as $(x_m, y_n)(x_m = (m-1)\Delta x, y_n = (n-1)\Delta y)(m = 1, \ldots M, n = 1, \ldots N)$, where $L_x = M \cdot \Delta x$, $L_y = N \cdot \Delta y$. The height can be given as

$$F(x_m, y_n) = \frac{1}{L_x L_y} \sum_{m_k=-M/2}^{M/2-1} \sum_{n_k=-N/2}^{N/2-1} F(k_{m_k}, k_{n_k}) exp[j(k_{m_k} x_m + k_{n_k} y_n)]$$

Where $F(k_{m_k}, k_{n_k})$ is the IFFT of PM spectrum. And then, the characteristic which is determined by height can be given as

$$\delta = \sqrt{\left(F(x)^2\right)_{ave} - F(x)^2_{ave}} \tag{5}$$

Besides, the root mean square height has the following conversion relationship with the PM spectrum:

$$\delta = \sqrt{\int S(k)dk} \tag{6}$$

Third, the reflection loss is calculated based on formula (9). For the reflection problem of rough ocean surface, we use formula (7) to estimate it. The attenuation factor is given by the International Radio Consultants Commission (CCIR) [2].

$$R^{'} = \rho R \tag{7}$$

$$\rho = \frac{1}{\sqrt{3.2g - 2 + \sqrt{(3.2g)^2 - 7g + 9}}} \tag{8}$$

$$L_{g0} = \rho E \tag{9}$$

where $g = 0.5\left(\frac{4\pi\delta f \sin\theta}{c}\right)^2$; c is light speed; f is frequency of electromagnetic waves; δ is sea level root mean square height; ρ is attenuation factor.

Finally, we add the reflection loss to transmission loss model and do the calculation.

3.2 Optimization Based on Searching Algorithm

After building the transmission loss model, which is shown in Fig. 2, it can be seen that electromagnetic waves pass through different distances in space under different jump times, which lead to different free space loss, ionosphere loss and reflection loss. If the

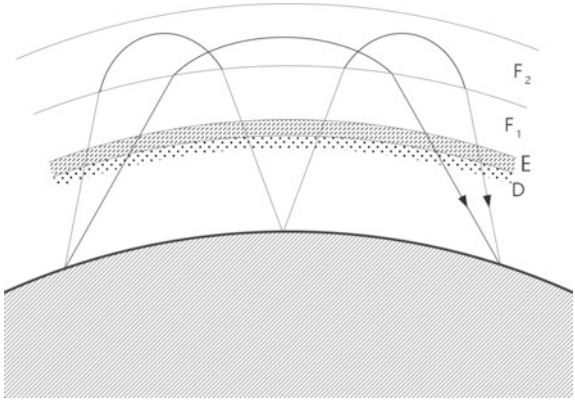


Fig. 2. Multi-hop situations (one hop, multiple hops)

signal hops many times, the loss will be accumulated and the signal will decay a lot. Therefore, we use the enumerative algorithm to find the optimal emission angle which is equivalent to finding the best SNR. The formula of searching algorithm is given by

$$SNR_{best} = max\{SNR_1, SNR_2, SNR_3 \ldots\} \tag{10}$$

Based on the above discussions, in order to evaluate the signal strength of receiver, the SNR is used to evaluate the signal strength of receiver as the communication quality standard. We regard the 10 dB of SNR as the threshold. As the propagation of noise is affected by ionosphere, its intensity varies with frequency, time, season, and so on. Due to the relatively complex calculation of atmospheric noise field strength and the focus of our discussion, we do not discuss the details of calculating atmospheric noise. We use the results from ITU P.372 [4].

$$E_{noise} = -0.0019f^4 + 0.098f^3 - 1.8f^2 + 13f - 13(\text{dB}) \tag{11}$$

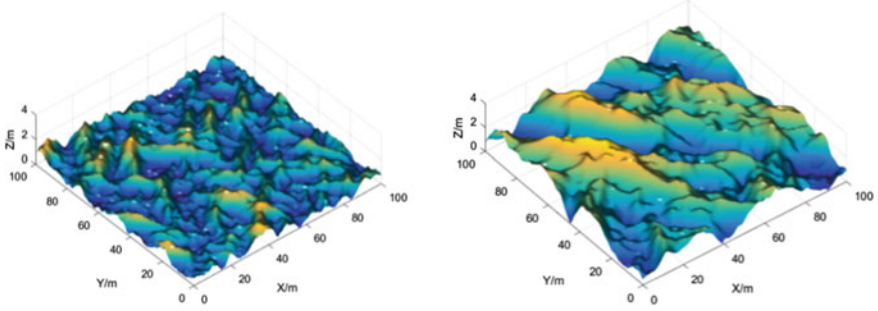
And SNR can be calculated based on the following formula:

$$SNR = E_n / E_{noise}(\text{dB}) \tag{12}$$

After calculating the SNR of different paths, we could get the best SNR based on formula (10).

4 Simulation Results

An example for the turbulent surface is shown in Fig. 3 when the launch angle is 45°. We can see the ocean surface with various wind speeds is formed with an altitude of 19.5 m above sea level. Therefore, the relationship between the wind speed and the elevation of ocean surface can easily be demonstrated in the following figures.



(a)Wind speed=3m/s (b)Wind speed=5m/s

Fig. 3. Ocean surface in different wind speed

The attenuation factor could be simulated based on formula (4), and the result proves that with the increase of wind speed, the attenuation factor increases. Also, we do some simulations with different frequencies and emission angles, and the results show the same tendency (Fig. 4).

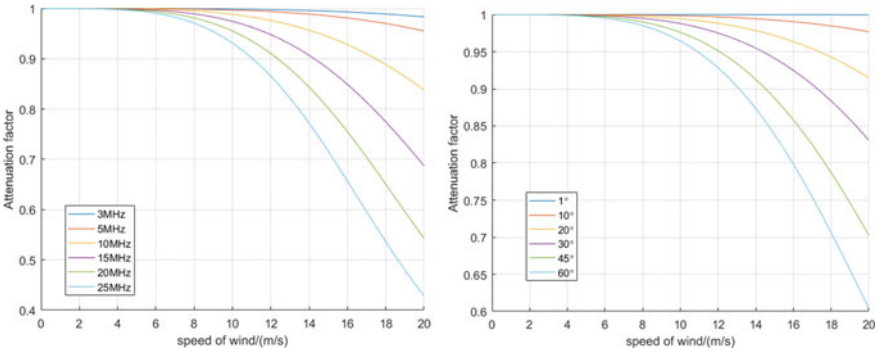


Fig. 4. Relationship between wind speed and attenuation factor

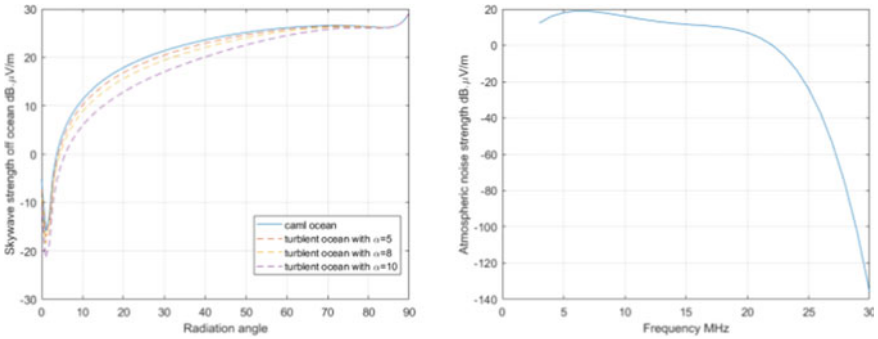
Then, since deriving the attenuation factor, both calm ocean's and turbulent ocean's transmission loss can be calculated. After calculating the transmission loss, we can calculate the field strength according to the formula (1).

The parameters which are needed for simulation and calculation are shown in Table 1.

Table 1. Related data in the calculation

Parameters	Values	Parameters	Values
P	0.1 kW	H	110 km
G	2	C	3×10^8 m/s
P	6730 km	ω	5 m/s, 8 m/s, 10 m/s
f	14.5 MHz	ε	70 + 90i
D	1000 km		

Figure 5(A) shows the wave strength with different emission angles, and we could get the result from Fig. 5(B) that the noise is weakened with the increase of frequency. Hoping times of different frequencies on the ocean are shown in Fig. 6. Afterward, it is easy to find the maximum hops under the required SNR (10 dB). In order to search for the maximum SNR, we plot the relationship between the SNR and emission angle. Based on Fig. 7, we could determine the best emission angle. In conclusion, the result shows the search for the highest SNR based on enumerative algorithm is effective, and it can help users set the optimal emission angle in transmission system.



(A) Comparison of electromagnetic field strength at different wind speeds over the sea surface
(B) The intensity of noise varies with the frequency of the electromagnetic wave

Fig. 5. Research about relationship between noise and signal strength

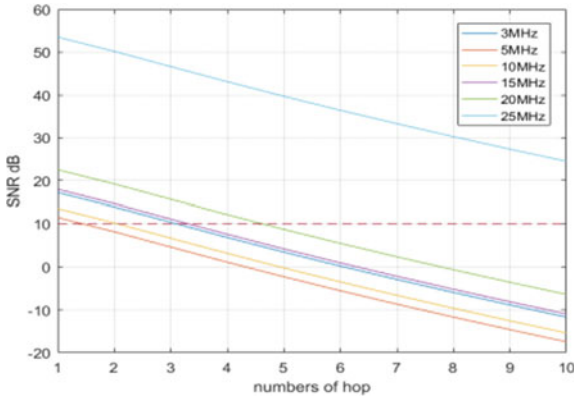


Fig. 6. Relationship between hopping times and SNR

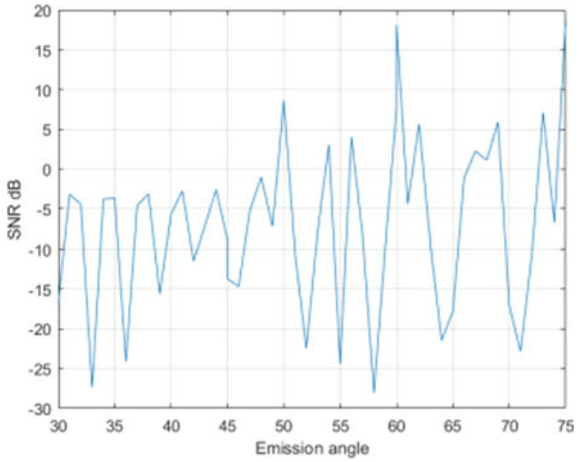


Fig. 7. Relationship between SNR and emission angle

5 Conclusion

In this paper, an optimal transmission path and an appropriate emission angle are taken into account by the algorithm based on the transmission loss model. Compared with the previous research, through simulating the dynamic change of sea surface, we calculate the attenuation effect of sea, which is used to calculate the reflection loss. Besides, we select the optimal emission angle based on the distance and signal-to-noise ratio. We provide some theoretical results of the long-range short-wave communication in ocean-going navigation. In the future, our research could be a pragmatic way to improve the quality of ship communication.

References

1. Ziemer, R.E., Tranter, W.H.: Principles of Communications, Systems, Modulation and Noise, pp. 304–406. Wiley (2008)
2. Proakis, Jhon: Digital Communications, 6th edn, pp. 200–203. Jhon Proakis, Perason (2009)
3. Tinin, M.V., Knizhin, S.I.: On radio wave propagation in multiscale randomly inhomogeneous ionosphere. In: Progress in Electromagnetics Research Symposium-Spring (PIERS), St. Petersburg, pp. 1278–1284 (2017). https://doi.org/10.1109/piers.2017.8261946
4. Barabashov, B.G., Anishin, M.M., Pelevin, O.Y.: High-frequency field strength prediction for ionospheric propagation at short- and medium-range radio paths. Radio Sci. **44**(01), 1–6 (2009). https://doi.org/10.1029/2008rs004038
5. Guan, Z., Chen, J., Bao, Z.: A scheme of auto ship detection for sky wave OTH radar within short CITs. In: IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Xiamen, pp. 1–4 (2017). https://doi.org/10.1109/icspcc.2017.8242367
6. Lott, G.K., Stark, W.R., Bail, M.T.: Multi-location long-term HF noise measurements and comparison to ITU-R P.372-8. In: International Conference on Ionospheric Radio Systems and Techniques, pp. 133–137. IEEE (2009)
7. Marchand, R.T., Brown, G.S.: Inferring rough surface parameters from average scattering data using approximate scattering models: 2. Pierson-Moskowitz spectrum. Radio Sci. **33**(4), 835–843 (1998). https://doi.org/10.1029/98rs01101
8. Zhang, Haiyong, Chi, Xu, Huang, Xiaofei: Evaluation of SNR and SIR based on HF field-strength prediction. Sci. Res. **13**, 852–854 (2011)

A New Robust Adaptive Beamforming Algorithm Based on Matrix Projection Method

Heping Shi[1], Zhiwei Guan[1(✉)], Lizhu Zhang[1], and Ning Ma[2]

[1] School of Automobile and Transportation, Tianjin University of Technology and Education (TUTE), Tianjin 300222, China
781004856@qq.com
[2] School of Electronic Engineering, Tianjin University of Technology and Education (TUTE), Tianjin 300222, China

Abstract. In order to avoid the desired signal cancellation, a new robust adaptive beamformer based on matrix projection method is developed. The proposed method mainly adopts spatial division method and Capon beamformer. The computation complexity of the proposed beamformer is in the same order as the complexity of Capon beamformer. Moreover, there is no iteration for the optimum solution in the proposed algorithm, which is more efficient in computation. Theoretical analysis and simulation results demonstrate that the proposed beamformer can effectively suppress the information of the desired signal.

Keywords: Robust beamforming · Adaptive beamforming · Interference-plus-noise space · Projection-based method · Desired signal cancellation

1 Introduction

The adaptive beamforming technology is an important research content in the field of array signal processing, which is widely used in such areas as wireless communication, underwater acoustic signal processing, medical imaging, space radio, radar signal processing, etc. [1–3]. However, the performance of the adaptive beamformer is apt to be effected by steering vector mismatch and formation error and some other unsatisfactory conditions [4]. When the desired signal existed in received data, the traditional adaptive beamformer would regard the desired signal as the interference and suppress it, whose performance would drop sharply, and this phenomenon is called desired signal cancellation. Therefore, the robust beamforming technology has become the research focus of the subject [5, 6], in which the minimum variance distortionless beam-forming technology attracts much attention. The literature [7] summarizes some robust design principles of MVDR beamforming technology such as sidelobe cancellation, diagonal loading, main space projection, etc.

The traditional beamforming based on projection method and the principal component beamforming belong to eigenspace based beamforming technology, and both are essentially the same. The projection beamforming algorithm projects the predicted desired signal steering vector into the signal-plus-interference subspace, and uses its

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 188–192, 2019.
https://doi.org/10.1007/978-981-13-6264-4_22

projection component as the presumed signal steering vector [8]. The principal component beamforming algorithm adopts the main space (and signal-plus-interference) of the data covariance matrix to perform MVDR beamforming.

2 Array Signal Models

Suppose there is a uniform linear array composed of M isotropic transducers, the array element spacing is half wavelength, the received signal is a far-field narrowband signal and the desired signal is uncorrelated with the interference. At the time of k, the data received by the array can be expressed as Eq. (1)

$$\mathbf{x}(k) = \mathbf{s}(k) + \mathbf{i}(k) + \mathbf{n}(k) \tag{1}$$

where $\mathbf{s}(k) = s(k)\mathbf{d}(\theta_s)$, $\mathbf{i}(k)$ and $\mathbf{n}(k)$, respectively, represent the statistic independent desired signal, interference and noise, $s(k)$ is the waveform of the desired signal, and $\mathbf{d}(\theta_s)$ is the steering vector of the desired signal. For the uniform linear array, an arbitrary angle steering vector is defined as Eq. (2)

$$\mathbf{d}(\theta) = \begin{bmatrix} 1 & e^{j\pi \sin\theta} & \cdots & e^{j\pi(M-1)\sin\theta} \end{bmatrix}^T \tag{2}$$

The theoretically received signal covariance matrix mathematical can be expressed as

$$\mathbf{R} = \sigma_s^2 \mathbf{d}(\theta_s)\mathbf{d}^H(\theta_s) + \sum_{i=1}^{Q} \sigma_i^2 \mathbf{d}(\theta_i)\mathbf{d}^H(\theta_i) + \sigma_n^2 \mathbf{I} \tag{3}$$

where $\left(\sigma_s^2, \{\sigma_i^2\}_{i=1}^{Q}\right)$, respectively, represent receiving $Q+1$ the energy of uncorrelated signals; $\left(\theta_s, \{\theta_i\}_{i=1}^{Q}\right)$ respectively represent the direction of signal arrival and θ_0 is the desired signal, $\{\theta_i\}_{i=1}^{Q}$ is Q interference signals. $\sigma_n^2\mathbf{I}$ represents noise covariance matrix and $\mathbf{I}$ is the identity matrix.

In fact, the theoretical covariance matrix cannot be obtained. In the simulation experiment of the thesis, the calculation adopts data sampling covariance matrix. The data covariance matrix obtained from N sampled data can be defined as follows:

$$\mathbf{R}_x = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_n(k)\mathbf{x}_n^H(k) \tag{4}$$

The output of adaptive beamforming is

$$y(k) = \mathbf{w}^H \mathbf{x}(k) \tag{5}$$

where $\mathbf{w} = [w_1, \ldots, w_M]^T$ is the complex weighted vector of the array.

The beamforming optimal output signal-to-interference-plus-noise ratio (SINR) can be achieved as follows:

$$\mathrm{SINR} = \frac{\sigma_s^2 |\mathbf{w}^H \mathbf{d}(\theta_0)|^2}{\mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w}} \tag{6}$$

where $\mathbf{R}_{i+n}$ represents the covariance matrix of the noise-plus-interference.

3 Spatial Division Technology

This section introduces the spatial division technology described in Literature [9].

Define an angular area Θ that needed to be divided. Establish the following matrix structure:

$$\mathbf{C} = \int_{\Theta} \mathbf{d}(\theta)\mathbf{d}^H(\theta)d\theta \tag{7}$$

Singular value decomposition for C, the corresponding left vectors group $\mathbf{U}_1$ of the first K large singular values and the rest constitute $\mathbf{U}_2$. Construct projection matrix structure Eq. (8) and Eq. (9), respectively.

$$\mathbf{P}_{\mathbf{U}_1} = \mathbf{U}_1 \mathbf{U}_1^H \tag{8}$$

$$\mathbf{P}_{\mathbf{U}_2} = \mathbf{U}_2 \mathbf{U}_2^H \tag{9}$$

The projection of the steering vector is defined as $\mathbf{P}_{U_1}\mathbf{d}(\theta)$ and $\mathbf{P}_{U_2}\mathbf{d}(\theta)$, respectively.

Figure 1 shows the effect of spatial division. The value is calculated from the squared norm of the vector obtained from the projection, where $M = 30$, $\Theta = [-5°, 5°]$ and $K = 5$.

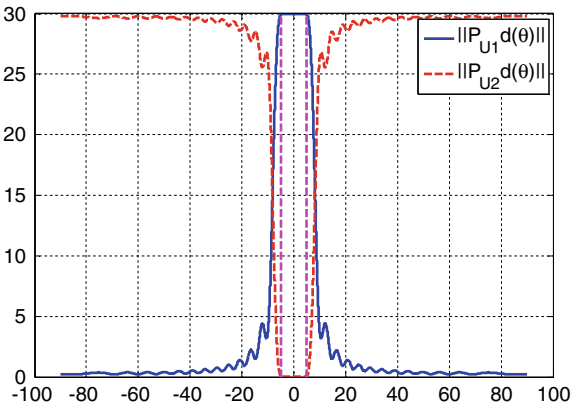


Fig. 1 The effect of spatial division

4 The Proposed Beamforming Algorithm Based on Matrix Projection

The projection matrix to divide signal space and interference-plus-noise space can be constructed by using the spatial division technology introduced in the previous section and combined with the known area where the incoming signal is located. The theoretical covariance matrix is first analyzed, and the Eq. (3) can be rewritten as follows:

$$\mathbf{R} = \int_{\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right)} \sigma^2(\theta)\mathbf{d}(\theta)\mathbf{d}^H(\theta)d\theta \tag{10}$$

Define the covariance matrix projection formula

$$\begin{aligned} \widehat{\mathbf{R}} &= \widehat{\mathbf{P}}_{\mathbf{U}_1} \cdot \mathbf{R} \cdot \widehat{\mathbf{P}}_{\mathbf{U}_1}^H \\ &= \int \sigma^2(\theta) \cdot \left(\widehat{\mathbf{P}}_{\mathbf{U}_1}\mathbf{d}(\theta)\right) \cdot \left(\widehat{\mathbf{P}}_{\mathbf{U}_1}\mathbf{d}(\theta)\right)^H d\theta \end{aligned} \tag{11}$$

where $\widehat{\mathbf{P}}_{\mathbf{U}_1}$ is constructed with the area where the desired signal is located. Equation (11) achieves the projection of the covariance matrix into the interference-plus-noise space. It can be seen that the information of the desired signal will be suppressed in the projected matrix. In the simulation experiment of the next section, the data covariance matrix replaces the theoretical covariance matrix for projection operations.

The experiment shows that the matrix obtained in Eq. (11) still needs to be diagonally loaded to recover to the normal noise level, and the loading is the noise average power.

$$\hat{\mathbf{R}} = \widehat{\mathbf{R}} + \hat{\sigma}_n^2\mathbf{I} \tag{12}$$

where $\hat{\sigma}_n^2$ is the estimation of noise power, which can be replaced by 10 times of the minimum eigenvalue of the data covariance matrix R.

Substituting $\hat{\mathbf{R}}$ into the Capon beamformer, the weight vector can be obtained.

5 Conclusion

The thesis proposes a beamforming technology based on the matrix projection. Construct the projection matrix of the interference-plus-noise space by using spatial division technology under the condition of the desired signal area is known, and then the data covariance matrix is projected into the interference-plus-noise space to suppress the desired signal energy, which effectively improves larger input signal-to-noise ratio and the cancellation of the desired signal when the steering vector mismatches. The computational complexity of the proposed algorithm is $O(M^3)$, the same as that of the Capon beamformer, and the proposed algorithm does not require iterative search of optimal solutions, which greatly saves computing resources and is easier to implement

in computers and hardware. The simulation results show that the proposed beamforming algorithm outperforms the comparison algorithms when the signal-to-interference-plus-noise ratio is relatively large under the mismatch conditions such as directivity error or non-correlated local scattering jamming.

Acknowledgment. This work was supported in part by the Natural Science Foundation of Tianjin under Grant No. 18JCQNJC01500, by the National Key Research and Development Program of China under Grant No. 2017YFB0102501, by the Artificial intelligence Science and Technology Support planning Major project of Tianjin under Grant No. 17ZXRGGX00070, by the Tianjin Municipal Science and Technology Innovation Platform, Intelligent Transportation Coordination Control Technology Service Platform under Grant No. 16PTGCCX00150 and by the Scientific Research Program of Tianjin Municipal Education Committee under Grant No. JWK1609.

References

1. Van Veen, B.D., Buckley, K.M.: Beamforming: a versatile approach to spatial filtering. IEEE ASSP Mag. **5**(2), 4–24 (1988)
2. Demir, Ö.T., Tuncer, T.E.: Antenna selection and hybrid beamforming for simultaneous wireless information and power transfer in multi-group multicasting systems. IEEE Trans. Wireless Commun. **15**(10), 6948–6962 (2016)
3. Zhang, Q., Zeng, X., Li, Q., Qin, J.: Joint beamforming and antenna subarray formation for nonregenerative MIMO relay networks. IEEE Trans. Veh. Technol. **65**(3), 1825–1830 (2016)
4. Vorobyov, S.A., Gershman, A.B., Luo, Z.Q.: Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem. IEEE Trans. Signal Process. **51**(2), 313–324 (2003)
5. Stoica, P., Wang, Z., Li, J.: Robust capon beamforming. IEEE Signal Process. Lett. **10**(6), 172–175 (2003)
6. Hassanien, A., Vorobyov, S.A., Wong, K.M.: Robust adaptive beamforming using sequential quadratic programming: an iterative solution to the mismatch problem. IEEE Signal Process. Lett. **15**, 733–736 (2008)
7. Vorobyov, S.A.: Principles of minimum variance robust adaptive beamforming design. Sig. Process. **93**, 3264–3277 (2013)
8. Chang, L., Yeh, C.C.: Performance of DMI and eigenspace-based beamformers. IEEE Trans. Antennas Propag. **40**(11), 1336–1347 (1992)
9. Khabbazibasmenj, A., Vorobyov, S.A., Hassanien, A.: Robust adaptive beamforming based on steering vector estimation with as little as possible prior information. IEEE Trans. Signal Process. **60**(6), 2974–2987 (2012)

Optimal Jamming Against OFDM/16QAM System

Hui Rong[1,2,3(✉)], Zhiguo Sun[2], Lili Guo[2], and Tianyu Xu[4]

[1] Harbin Engineering University, Nantong Street 145, Harbin 150001, China
ronghui@hrbeu.edu.cn
[2] College of Information and Communication Engineering, Harbin Engineering University, Harbin, China
[3] Department of Electrical and Computer Engineering, University of California, San Diego, USA
[4] Huawei Technologies Co., Shanghai, China

Abstract. Bit error rate (BER) of Orthogonal Frequency Division Multiplexing (OFDM) corrupted by traditional jamming has already been studied. In order to save jamming power and figure out the optimal jamming against OFDM, BER of OFDM corrupted by two-dimensional jamming is investigated. In this paper, the accurate BER formula of MQAM and OFDM/MQAM under two-dimensional jamming and AWGN is derived. Based on the BER formula, optimal jamming against 16QAM and OFDM/16QAM was taken as an example for analyzing. Simulation results verified that it is not always optimal to match the jamming signal to the victim signal. Theory derivation and simulation all verify the result that OFDM/QPSK is the optimal jamming against OFDM/16QAM.

Keywords: OFDM · Optimal jamming · 16QAM · Convex analysis

1 Introduction

Orthogonal Frequency Division Multiplexing (OFDM) is a promising technology which is robust against multipath fading and inter-symbol interference (ISI). OFDM has been widely used in lots of area such as public video and voice broadcasting. In addition, OFDM is the main technique in air-interface of Long-Term Evolution (LTE). Orthogonal Frequency Division Multiplexing Access (OFDMA) which is based on OFDM will be applied to the air-interface of 5G which is now heated discussing.

For the importance of OFDM in wireless communication, it is urgent to figure out the optimal jamming against OFDM for the future electronic warfare. Earlier literatures have studied the influence of traditional jamming against OFDM system. The influence of barrage jamming against OFDM systems are introduced in [1–4]. Due to the low efficiency of barrage jamming, partial barrage jamming is discussed in [3–5]. The impact of noise jamming against OFDM system is investigated in [5, 6]. Uncorrelated jamming and narrow band jamming against OFDM are mentioned in paper [7, 8]. Besides, OFDM is a multi-carrier modulation which can shut down some of its sub-carriers to negative the influence of single-tone, multi-tone, and narrow band jamming.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 193–200, 2019.
https://doi.org/10.1007/978-981-13-6264-4_23

A noteworthy mention would be [3], where Lou et al. summarize the bit error rate (BER) of OFDM with traditional jamming.

As modern communication reconnaissance technique has strong ability in parameter detection of noncooperative communication systems, communication reconnaissance-based optimal jamming is of great practical value. Earlier literatures consider that the optimal jamming must match the victim's parameters. This paper shows that it is not always optimal to match the jammer's modulation scheme and parameters to the victim's in order to maximize the BER.

The rest of the paper is organized as follows: The OFDM system is introduced in Sect. 2. The accurate BER formula of Quadrature Amplitude Modulation (M-QAM) corrupted by two-dimensional modulation jamming and additive white Gaussian noise (AWGN) is derived in Sect. 3. Particularly, optimal jamming modulation scheme against 16-QAM and OFDM/16-QAM is analyzed by the method of convex analysis. In Sect. 4, simulation results and numerical analysis are presented. Finally, this paper is concluded in Sect. 5.

2 OFDM/16QAM System Model

The overview of OFDM system model based on IFFT and FFT is illustrated in Fig. 1. 16-QAM is used as the sub-carriers' modulation scheme. OFDM symbols are transmitted under AWGN channel.

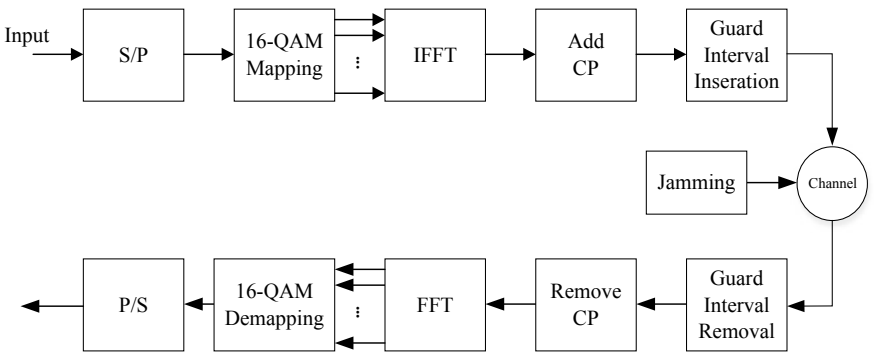


Fig. 1. System diagram of OFDM

OFDM symbol is expressed by

$$s(l) = \sum_{k=1}^{N_{sc}-1} S(k) \exp\left(\frac{\mathrm{j}2\pi lk}{N_{sc}}\right), \ \forall 0 \le l \le N_{sc} - 1 \tag{1}$$

where S(k)are the 16-QAM modulated symbols and N_{sc} is the total number of sub-carriers used in OFDM.

Assume that OFDM signal is transmitted under AWGN channel, the received signal at the input end of 16-QAM de-mapper can be represented as

$$Y(k) = \sum_{l=0}^{N_{sc}-1} \left\{ \left[\sqrt{P_s} s(l) + n(l) \right] \exp\left(\frac{-\mathrm{j}2\pi kl}{N_{sc}} \right) \right\}, \ \forall 0 \le k \le N_{sc} - 1 \tag{2}$$

where $n(l)$ is additive white Gaussian noise, P_s is the average received power.

For the reason that the characters of $n(l)$ is the same as the FFT of $n(l)$, OFDM/16QAM system performs no differently from the 16-QAM system under AWGN [9].

Now we discuss the scenario where jamming exists. Suppose a coherent receiver and perfect synchronization, the received signal corrupted by jamming signal can be expressed as

$$y(l) = \left[\sqrt{P_s} s(l) + \sqrt{P_j} j(l) \right] \exp(\mathrm{i}\omega_0 t) + n(l) \tag{3}$$

where $j(l)$ is the jamming symbols, P_j is the average jamming power.

Since the jamming signal discussed in this paper are all considered as two-dimensional jamming. Namely,$\bar{y}_m = \bar{y}(l = mT) = [\Re y_m, \Im y_m]^T$, where $\Re y_m$ is the real part of y_m, $\Im y_m$ is the image part of y_m. Thus, the received signal can be rewritten as

$$\bar{y}_m = \bar{y}(l = mT) = (\sqrt{P_s}\bar{s}_m + \sqrt{P_j}\bar{j}_m) \exp(\mathrm{i}\omega_0 t) + \bar{n}_m, \ m = 1, 2 \ldots \tag{4}$$

where T is the symbol interval.

3 BER of MQAM with Jamming and Convex Analysis

This section is our main technical part. We derive the accurate BER formula of M-QAM with jamming for the conclusion that OFDM/16-QAM performs no differently from the 16-QAM system. Optimal jamming modulation scheme based on the BER formula against 16-QAM and OFDM/16-QAM system is analyzed by using the convex analysis.

3.1 BER of M-QAM with Two-Dimensional Jamming

Note that the constellation of rectangular M-QAM can be considered as two $\sqrt{\mathrm{M}}-PAM$ constellations in the in-phase and quadrature dimensions [10]. The minimum distance between any two points is $d_{\min}$ which is given by

$$d_{\min} = \sqrt{\frac{12 \log_2 M}{M^2 - 1} E_b} \tag{5}$$

where E_b is the average bit energy, M is the modulation order. The constellation points of M-PAM are uniformly located at $\{\pm 0.5 d_{\min}, \pm 1.5 d_{\min}, \ldots, \pm 0.5(M-1) d_{\min}\}$.

The BER of M-QAM without jamming is given by

$$P_e = (1 - \frac{1}{M})\text{erfc}\left(\sqrt{\frac{3\log_2 M}{M^2 - 1}\frac{E_b}{N_0}}\right) \tag{6}$$

Suppose that the jammer has the knowledge of victim signals which is obtained by communication reconnaissance such as symbol interval and carrier frequency, etc. Thus, the received signal in either dimension can be expressed as

$$\begin{aligned} x(t) &= 2\int_0^T \bar{y}(t)\cos(\omega_0 t)dt \\ &= \underbrace{2\int_0^T \sqrt{P_s}\bar{s}_m\hat{g}(t)\cos^2(\omega_0 t)dt}_{S(t)} + \underbrace{2\int_0^T \sqrt{P_j}\bar{j}_m\hat{g}(t)\cos^2(\omega_0 t)dt}_{J(t)} + \underbrace{2\int_0^T \bar{n}_m\cos(\omega_0 t)dt}_{N(t)} \end{aligned} \tag{7}$$

where $g(t)$ is the pulse shape, $\hat{g}(t)$ is the output of the match filter given by $\hat{g}(t) = g(t) * g(t)$.

For convenience, $g(t)$of victim signal and jamming signal are all considered as rectangular pulse. Since $N(t)$is subjected to $N(t) \sim N(0, N_0)$ in a symbol interval, the mean and the variance of the decision value is $J(t)$ and N_0, respectively. The probability density function (PDF) of decision value can be represented as

$$f(x) = \exp\left\{-[x - J(t)]^2/(2N_0)\right\}/\sqrt{2\pi N_0} \tag{8}$$

When the symbols "0" and "1" is sent with equal probability, the BER is derived as

$$P_e = \frac{1}{4}\left[\text{erfc}\left(\sqrt{\frac{3\log_2 M}{M^2 - 1}}\sqrt{\frac{E_b}{N_0}} + \frac{J(t)}{\sqrt{2N_0}}\right) + \text{erfc}\left(\sqrt{\frac{3\log_2 M}{M^2 - 1}}\sqrt{\frac{E_b}{N_0}} - \frac{J(t)}{\sqrt{2N_0}}\right)\right] \tag{9}$$

Let us denote the ratio of jamming average bit energy E_j to noise power density N_0 by E_j/N_0. For the reason that all jamming symbols can be separated into two dimension, namely $jam_n = \Re\bar{j}_m$ *or* $\Im\bar{j}_m$. Without loss of generality, formula (9) can be rewritten as

$$P_e = \frac{1}{l}\sum_{n=1}^{l}\left\{\frac{1}{4}\left[\text{erfc}\left(\sqrt{\frac{3\log_2 M}{M^2 - 1}}\sqrt{\frac{E_b}{N_0}} + \sqrt{\frac{E_j}{N_0}}jam_n\right) + \text{erfc}\left(\sqrt{\frac{3\log_2 M}{M^2 - 1}}\sqrt{\frac{E_b}{N_0}} - \sqrt{\frac{E_j}{N_0}}jam_n\right)\right]\right\} \tag{10}$$

where l is the total number of constellation points on either dimension, namely, $n = 1, 2, 3 \ldots l$. As the two dimensions are orthogonal, the total BER P_{et} of 16-QAM system is given by

$$P_{et} = 1 - (1 - P_e)^2 \tag{11}$$

3.2 Optimal Jamming Against 16QAM

Let us denote Eq. (10) by $P_e(jam_n, SNR, JNR)$ for convenience. Obviously, P_e is symmetric in jam_n. Thus,

$$P_e(jam_n, SNR, JNR) = P_e(-jam_n, SNR, JNR) \tag{12}$$

Therefore, P_e is maximized over the distribution of $a = |jam_n|$ which means transmit $jam_n = a$ and $jam_n = -a$ with equal probability. As discussed above, the optimization problem can be described as

$$\max_{f_J} \int_a P_e(a, SNR, JNR) f_J(a) da \equiv \max_{f_J} E(P_e(a, SNR, JNR)) \tag{13}$$

In order to solve the above problem, we should first derive how to express the jamming signal PDF $f_J(a)$.

First, the following set is defined:

$$U = \{(u_1, u_2) : u_1 = P_e(a, SNR, JNR), u_2 = a^2\}$$
$$W = \{(w_1, w_2) : w_1 = E_{f_J}(P_e(a, SNR, JNR)), w_2 = E_{f_J}(a^2)\}$$

Let V represent the convex hull of U. If we want to prove that $W = V$, we need to prove $V \subseteq W$ and $W \subseteq V$, respectively.

(1) $V \subseteq W$

Since V is the convex hull of U, each element of V can be expressed as follows:

$$\upsilon = \sum_{i=1}^{l} \lambda_i (P_e(a, SNR, JNR), a^2) \tag{14}$$

where $\sum_{i=1}^{l} \lambda_i = 1, \lambda_i \geq 0, \forall i$. Considering set W, it has an element that is equal to υ if $f_J(a) = \sum_{i=1}^{l} \lambda_i \delta(x - x_i)$. Hence, each element of V all exists in W.

(2) $W \subseteq V$

Paper [11] has proved that for any vector random variable Θ that takes values in set Ω, the expected value $E\{\Theta\}$ of Θ is in the convex hull of Ω. That is $W \subseteq V$.

Since $V \subseteq W$ and $W \subseteq V$, it is concluded that $W = V$.

Therefore, Carathéodory theorem implies that any point in V can be expressed as a convex combination of at most three points in U. Since the optimal jamming signal PDF should maximize the objective function, the optimal solution exists on the boundary. Since any point on the boundary can be expressed as a convex combination of at most 2 elements in U, the optimal jamming signal PDF $f_J(a)$ is given by

$$f_J(a) = \lambda\delta(a - a_1) + (1 - \lambda)\delta(a - a_2),\ \lambda \in (0, 1) \tag{15}$$

where $\lambda a_1^2 + (1 - \lambda)a_2^2 \leq 1$, λ and $1 - \lambda$ are the probabilities with which the jammer sends signals a_1 and a_2, respectively along any dimension.

Obviously, $P_e(a, SNR, JNR)$ is a nondecreasing function that will be maximized on the boundary.

Let us define four variables

$$x_1 = \alpha\sqrt{SNR} + \sqrt{JNR}a_1, x_2 = \alpha\sqrt{SNR} - \sqrt{JNR}a_1$$
$$y_1 = \alpha\sqrt{SNR} + \sqrt{JNR}a_2, y_2 = \alpha\sqrt{SNR} - \sqrt{JNR}a_2$$

where $\alpha = \sqrt{(3\log_2 M)/(M^2 - 1)}$.

Therefore, the overall P_e can be rewritten as

$$P_e(\lambda, a_1, a_2, SNR, JNR) = \frac{k}{l}\sum_{n=1}^{l}\{\lambda[\mathrm{erfc}(x_1) + \mathrm{erfc}(x_2)] + (1 - \lambda)[\mathrm{erfc}(y_1) + \mathrm{erfc}(y_2)]\} \tag{16}$$

When at the boundary, namely, $\lambda a_1^2 + (1 - \lambda)a_2^2 = 0.5$, a_2 can be represented as the function of a_1. The first-order partial differential of P_e over a_1 can be expressed as

$$\frac{\partial P_e}{\partial a_1} = \frac{k}{l}\sum_{n=1}^{l}\left\{-\frac{2\lambda}{\sqrt{\pi}}\sqrt{JNR}\left[e^{-x_1^2} - e^{-x_2^2}\right] + \frac{2\lambda}{\sqrt{\pi}}\sqrt{JNR}\frac{a_1}{a_2}\left[e^{-y_1^2} - e^{-y_2^2}\right]\right\} \tag{17}$$

$\partial P_e(\lambda, a_1, SNR, JNR)/\partial a_1 = 0$ is $a_1 = 0$ *or* $1/\sqrt{2}$, $\partial^2 P_e/\partial a_1^2 < 0$ *if* $a_1 = 1/\sqrt{2}$ will be achieved when $\sqrt{E_j/E_b} > d_{\min}^2/2$. For the first-order partial differential of P_e over a_1 is 0 and the second-order partial differential of P_e over a_1 is less than 0, P_e is maximized at $a_1 = 1/\sqrt{2}$. It is well known that the entropy of any two-point distribution is maximized when each of the points is equally likely. Therefore the jammer sends $a_1 = 1/\sqrt{2}$ with probability 0.5 and $a_1 = 1/\sqrt{2}$ with probability 0.5 either along in-phase and quadrature dimension respectively. Therefore the optimal jamming signal is QPSK against 16QAM.

3.3 Optimal Jamming Against OFDM/16QAM

After guard interval removal, CP removal and FFT, the received signal can be expressed as

$$
\begin{aligned}
Y(k) &= \sum_{l=0}^{N_{sc}-1} \left\{ \left[\sqrt{P_s} s(l) + \sqrt{P_j} j(l) + n(l) \right] \exp\left(\frac{-\mathrm{j}2\pi kl}{N_{sc}} \right) \right\} \\
&= \sqrt{P_s} S(k) + \sqrt{P_j} \sum_{l=0}^{N_{sc}-1} \left[j(l) \exp\left(\frac{-\mathrm{j}2\pi kl}{N_{sc}} \right) \right] + N(k)
\end{aligned} \tag{18}
$$

where $\forall 0 \le k \le N_{sc} - 1$, $N(k)$ is the FFT of $n(l)$.

Obviously, $S(k)$ is the 16-QAM modulated victim symbols. Based on the above analysis, optimal jamming signal against $\sum_{l=0}^{N_{sc}-1} j(l) \exp(-\mathrm{j}2\pi kl/N_{sc})$ is QPSK signal. Therefore, this leads to the conclusion that $j(l)$ is the IFFT of QPSK, or to say the optimal jamming signal against OFDM/16QAM is OFDM/QPSK signal.

4 Simulation Results

The simulation is divided into two parts. First, simulation results are compared with the theoretical results when using 16QAM modulated symbols as the victim signal to confirm the correctness of the derived BER formula of MQAM. Secondly, simulation has been done to verify the optimal jamming against OFDM/16QAM victim signal is OFDM/QPSK signal.

The BER of 16QAM jammed by BPSK, QPSK, and 16QAM is simulated respectively in Fig. 2. Particularly, it is assumed that the interval of a jamming symbol is same as that of the victim signal. Without doubt, the carrier frequency is the same as well. It is also assumed that $E_b/E_j = 10$ dB.

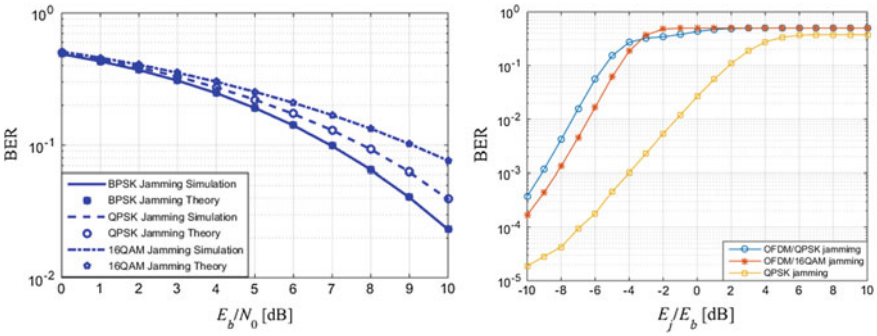


Fig. 2. BER of 16-QAM jammed by three kinds of modulation jamming and OFDM/16QAM jammed by OFDM/16QAM, OFDM/QPSK and QPSK jamming

Obviously, the simulation results agree with the theoretical results very well which verifies the correctness of the derived BER formula.

The performance of OFDM/QPSK modulated jamming and QPSK modulated jamming against OFDM/16-QAM victim signal is simulated in Fig. 2.

As results above, OFDM/QPSK jamming does well than OFDM/16-QAM jamming at $\sqrt{E_j/E_b} > -3$ dB which meet the theory.

5 Conclusions

This paper proposed the BER formula for MQAM signal corrupted by two-dimensional modulation jamming and AWGN. Optimal jamming against OFDM/16QAM victim signal is OFDM/QPSK signal but not OFDM/16QAM. The simulation results illustrate the efficiency of optimal jamming.

Acknowledgement. This paper is funded by the International Exchange Program of Harbin Engineering University for Innovation-oriented Talents Cultivation, the Fundamental Research Funds for Interference modeling and management in cognitive radio net-works (Grant No.61401196). The support provided by China Scholarship Council(CSC) during a visit of Hui Rong to University of California San Diego is acknowledged.

References

1. Rahimzadeh, P., Ashtiani, F.: Analytical evaluation of saturation throughput of a cognitive WLAN overlaid on a time-scheduled OFDMA network. IEEE Trans. Mob. Comput. **16**(3), 634–647 (2017)
2. Clancy, T.C.: Efficient OFDM denial: Pilot jamming and pilot nulling. In: Proceeding on IEEE International Conference Communication, pp. 1–5 (2011)
3. Luo, J., Andrian, J., Zhou, C.: Bit error rate analysis of jamming for OFDM systems. In: Proceedings of WTS, pp. 1–8 (2007)
4. Chi, D.W., Das, P.: Effects of jammer and nonlinear amplifiers in MIMO-OFDM with application to 802.11n WLAN. In Proceeding on IEEE Military Communication Conference, pp. 1–8 (2008)
5. Park, J., Kim, D., Kang, C., Hong, D.: Effect of partial band jamming on OFDM-based WLAN in 802.11 g. In: Proceedings of IEEE International Conference on Acoustics, Speech, Signal Processing, vol. 4, pp. 560–563 (2003)
6. Jha, R., Limkar, S., Dalal, U.: Performanceanalysisundertheinfluence of jamming for WiMAX system. In Proceedings of 2nd International Conference on Emerging Applications Information Technology, pp. 292–297 (2011)
7. Pollet, T., Bladel, M.V., Moeneclaey, M.: BER sensitivity of OFDM systems to carrier frequency offset and wiener phase noise. IEEE Trans. Commun. **43**(2–4), 191–193 (1995)
8. Marey, M., Steendam, H.: Analysis of the narrowband interference effect on OFDM timing synchronization. IEEE Trans. Signal Process. **55**(9), 4558–4566 (2007)
9. Fazel, K., Kaiser, S.: Multi-Carrier and Spread Spectrum Systems: From OFDM and MC-CDMA to LTE and WiMAX. Wiley, Hoboken, NJ, USA (2008)
10. Proakis, J.G., Salehi, M.: Digital Communications, 5th edn. McGraw-Hill, New York, NY, USA (2008)
11. Schilling, D.L., Milstein, L.B., Pickholtz, R.L.: Optimization of the processing gain of and M-ary direct sequence spread spectrum communication system. IEEE Trans. Commun. **28** (8), 1389–1398 (1980)

Low-Complexity Robust Adaptive Beamforming Based on Covariance Matrix Reconstruction

Zhiwei Guan[1], Heping Shi[1(✉)], Lizhu Zhang[1], and Ning Ma[2]

[1] School of Automobile and Transportation, Tianjin University of Technology and Education (TUTE), Tianjin 300222, China
shiheping@tju.edu.cn
[2] School of Electronic Engineering, Tianjin University of Technology and Education (TUTE), Tianjin 300222, China

Abstract. A new beamformer based on covariance matrix reconstruction is introduced. The essence of the new approach is to eliminate the desired signal component in the sample covariance matrix and thus complex integral operation is avoided in the procession of covariance matrix reconstruction. Besides, the actual array steering vector is estimated by a new technique. Contrary to other reference beamformers, simulation results demonstrate the effectiveness of our proposed method.

Keywords: Robust adaptive beamforming · Covariance matrix reconstruction · Steering vector estimation

1 Introduction

Adaptive beamforming has been one of the most important research areas in sensor array signal processing for decades, which has possessed many applications such as wireless communication and microphone array speech processing. It is well known that conventional adaptive beamforming such as standard Capon beamforming needs to assume exact knowledge of desired signal (DS). It will suffer significant performance degradation when the information of DS is not known accurately.

Numerous robust adaptive beamforming (RAB) techniques have been proposed to deal with this problem [1, 2]. For instance, diagonal loading is known to be a popular approach, however, it is difficult to choose the optimal loading factor. In [3], a new beamformer named worst-case performance optimization based on the second-order cone programming (SOCP) problem was reported. However, to solve this problem, specific optimization toolbox such as CVX is needed.

In order to achieve robustness against large look direction mismatch, a new method is introduced in this paper. In [4], the interference-plus-noise (IPN) covariance matrix was reconstructed by using the Capon spectrum to integrate over an angular sector separated from the direction of the desired signal. However, this caused complex integral operation. To reduce the computational complexity, the IPN covariance matrix is reconstructed by utilizing a new approach in this paper. The essence of the new

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 201–206, 2019.
https://doi.org/10.1007/978-981-13-6264-4_24

approach is to wipe up the desired signal component in the sample covariance matrix and thus complex integral operation can be avoided in the procession of covariance matrix reconstruction. Furthermore, unlike traditional uncertainty set based robust beamformers, it is not necessary to use a large size of the uncertainty set against direction-of-arrival (DOA) mismatch in the proposed algorithm.

2 Array Signal Models

Consider a uniform linear antenna array with N omnidirectional equispaced antenna sensor elements receiving multiple far-field narrowband signals. Let $\mathbf{n}(k)$ be a N*1 complex vector representing the array observation at the kth snapshot

$$\mathbf{x}(k) = \mathbf{s}(k) + \mathbf{i}(k) + \mathbf{n}(k) \tag{1}$$

where $\mathbf{s}(k) = s(k)\mathbf{d}(\theta_s)$, $\mathbf{i}(k)$ and $\mathbf{n}(k)$ respectively represent the statistic independent desired signal, interference and noise.

The beamformer output is given by

$$y(k) = \mathbf{w}^H\mathbf{x}(k) \tag{2}$$

where $\mathbf{w} = [w_1, \ldots, w_N]^T$ is the complex weight vector of the antenna array, and $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and Hermitian transpose operator.

In general, we introduce the output signal-to-interference-plus-noise ratio (SINR) to measure the beamformer performance. Assuming the desired signal steering vector is $\mathbf{a}(\theta_0)$, θ_0 is the DOA of the DS.

$$SINR = \frac{\sigma_s^2|\mathbf{w}^H\mathbf{a}(\theta_0)|^2}{\mathbf{w}^H\mathbf{R}_{i+n}\mathbf{w}} \tag{3}$$

where σ_s^2 denotes the DS power and the N*N ideal IPN covariance matrix $\mathbf{R}_{i+n}$ can be acquired from $\mathbf{R}_{i+n} = E\{[\mathbf{i}(k) + \mathbf{n}(k)][\mathbf{i}(k) + \mathbf{n}(k)]^H\}$.

The standard Capon beamformer can be obtained by solving the following optimization problem:

$$\min_{\mathbf{w}} \quad \mathbf{w}^H\mathbf{R}_{i+n}\mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^H\mathbf{a}(\theta_0) = 1 \tag{4}$$

The solution to (4) optimization problem is

$$\mathbf{w} = \frac{\mathbf{R}_{i+n}^{-1}\mathbf{a}(\theta_0)}{\mathbf{a}^H(\theta_0)\mathbf{R}_{i+n}^{-1}\mathbf{a}(\theta_0)} \tag{5}$$

Note that the perfectly known array covariance matrix $\mathbf{R}_{i+n}$ is unavailable and it is often replaced by the sample covariance matrix

$$\mathbf{R}_x = (1/K) \sum_{k=1}^{K} \mathbf{x}(k)\mathbf{x}(k)^H \tag{6}$$

K is the number of snapshots. As K increases, the matrix $\mathbf{R}_x$ will converge to the theoretical covariance matrix $\mathbf{R} = \sigma^2\mathbf{a}(\theta_0)\mathbf{a}(\theta_0)^H + \mathbf{R}_{i+n}$, that is to say, the desired signal steering vector lies in the sample covariance matrix $\mathbf{R}_x$. The DS may be treated as an interference signal, which leads to DS "self-cancelation", and thus the array system performance is affected. Therefore, it is necessary to eliminate the DS component in the sample covariance matrix $\mathbf{R}_x$.

3 The Proposed Algorithm

In the proposed method, instead of using the spatial distribution, the IPN covariance matrix is reconstructed by regulating sample covariance matrix, and the desired signal steering vector is estimated with the help of the parallelogram rule of vectors.

3.1 IPN Covariance Matrix Reconstruction

The essence of the proposed IPN covariance matrix reconstruction technique is to wipe out the desired signal component in the sample covariance matrix. To achieve this goal, the sample covariance matrix $\mathbf{R}_x$ is decomposed as follows:

$$\mathbf{R}_x = \sum_{i=1}^{N} \lambda_i \mathbf{e}_i \mathbf{e}_i^H = \mathbf{E}_s \mathbf{\Lambda}_s \mathbf{E}_s^H + \mathbf{E}_n \mathbf{\Lambda}_n \mathbf{E}_n^H \tag{7}$$

where $\{\lambda_i, i = 1, \ldots, N\}$ is the eigenvalues of matrix $\mathbf{R}_x$ in descending order, $\mathbf{e}_i$ is the eigenvector associated with λ_i. $\mathbf{E}_s = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{M+1}]$ and $\mathbf{E}_n = [\mathbf{e}_{M+2}, \ldots, \mathbf{e}_N]$. $\mathbf{\Lambda}_s = diag\{\lambda_1, \lambda_2, \ldots, \lambda_{M+1}\}$ and $\mathbf{\Lambda}_n = diag\{\lambda_{M+2}, \ldots, \lambda_N\}$ is a diagonal matrix. M is the number of interference.

As mentioned above, the DS component in the sample covariance matrix will reduce the beamformer performance. To minimize the impact of this component, refer to the idea of diagonal loading, the signal-plus-interference subspace matrix can be processed as follows:

$$\hat{\Lambda}_s = [\lambda_1 - 10\sigma_n^2, \lambda_2 - 10\sigma_n^2, \ldots, \lambda_{M+1} - 10\sigma_n^2] \tag{8}$$

However, nothing is done with other eigenvalues which are associated with the noise subspace. That is to say, all the eigenvalues associated with the signal-plus-interference subspace minus a small amount and the eigenvalues associated with the noise subspace remain the original sample. Although the above processing will also have an effect on the interference signal, this operation can better reduce the impact of

the DS component. The presence of the DS is critical to the performance of beamformer. Next, we recombine the adjusted matrix and get the reconstructed IPN covariance matrix.

$$\mathbf{R}_{re} = \mathbf{E}_s \hat{\Lambda}_s \mathbf{E}_s^H + \mathbf{E}_n \mathbf{\Lambda}_n \mathbf{E}_n^H \tag{9}$$

3.2 Desired Signal Steering Vector Estimation

In practical applications, for the actual DS steering vector, the presumed DS steering vector $\mathbf{a}(\theta_0)$ always exists mismatch. In [5], $\sqrt{N}\mathbf{e}_t$ is used as the estimation of the actual DS steering vector, where the eigenvalue vector $\mathbf{e}_t$ is most similar to the DS steering vector $\mathbf{a}(\theta_0)$. Simulation results have demonstrated that this estimation is also deviated from the actual value to a certain extent. In this paper, the actual DS steering vector is estimated as follows:

$$\widehat{\mathbf{a}} = \sqrt{N}\mathbf{e}_t + \mathbf{\Delta} \tag{10}$$

where $\mathbf{\Delta}$ denotes the mismatch vector with an expression of $\mathbf{\Delta} = \mathbf{a}(\theta_0) - \mathbf{e}_t$. That is to say, the actual DS steering vector can be expressed as the difference between the presumed DS steering vector $\mathbf{a}(\theta_0)$ and estimation of the actual DS steering in [3].

From a complexity point of view, the main computational cost in this algorithm is the eigendecomposition operation on the sample covariance matrix $\mathbf{R}_x$, which has a complexity of $O(N^3)$. Compared to the other RAB methods, the presented method does not need an integral operation or any optimization toolbox, so it can be applied to practical engineering easily.

4 Simulation Results

Computer simulations are performed using a 10-element uniform linear array with elements spaced a half wavelength apart. Three independent narrowband signal sources are present in the directions of 0°, 30°, and 50°, respectively. The first signal is the desired signal, and the other two are interferences. The noise is presumed to be white noise with unit covariance.

Three other RAB approaches are compared with the proposed method in terms of the array output signal-to-interference-plus-noise ratio (SINR): the traditional diagonal loading beamformer; the beamformer using worst-case performance optimization (WCB) [3], and the beamformer based on steering vector estimation with as little as possible prior information (API) [4]. The angular location of the signal of interest is presumed to be $\Theta = [\theta_0 - 5^\circ, \theta_0 + 5^\circ]$ in the beamformer of [3]. For each scenario, 100 independent Monte Carlo trials are used.

In the first example, we set $\theta_0 = 2^\circ$, that is to say, there is a 2° look direction error. In this example, the SNR equals 10 dB. The output SINR against the number of snapshots is investigated and the corresponding results are shown in Fig. 1.

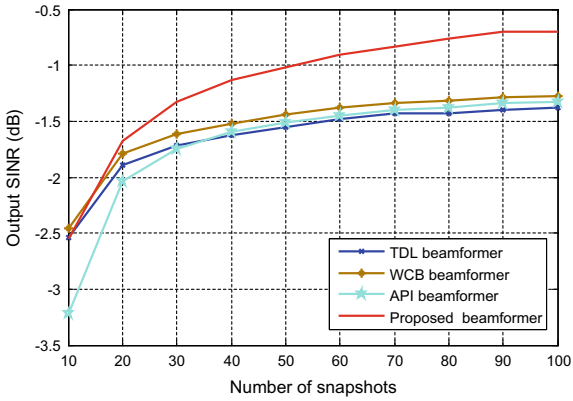


Fig. 1. Output SINR of beamformers against the number of snapshots

It is found that the proposed approach enjoys the best performance among the RAB methods tested and converges to a relatively high level with fewer snapshots. The performance improvement is a direct result of the DS elimination and the ASV estimation.

5 Conclusion

A new low-complexity RAB approach is presented in this letter. Based on the sample covariance matrix, in the proposed method, the IPN covariance matrix is reconstructed and the actual ASV is estimated using a desired signal steering vector and the eigenvectors of the sample covariance matrix. In contrast to other RAB approaches, the proposed method does not involve any optimization and integral operation, and it can achieve better performance.

Acknowledgment. This work was supported in part by the Natural Science Foundation of Tianjin under Grant No. 18JCQNJC01500, by the National Key Research and Development Program of China under Grant No. 2017YFB0102501, by the Artificial intelligence Science and Technology Support planning Major project of Tianjin under Grant No. 17ZXRGGX00070, by the Tianjin Municipal Science and Technology Innovation Platform, Intelligent Transportation Coordination Control Technology Service Platform under Grant No. 16PTGCCX00150 and by the Scientific Research Program of Tianjin Municipal Education Committee under Grant No. JWK1609.

References

1. Hassanien, A., Vorobyov, S.A., Wong, K.M.: Robust adaptive beamforming using sequential programming: an iterative solution to the mismatch problem. IEEE Signal Process. Lett. **15**, 733–736 (2008)
2. Gu, Y.J., Leshem, A.: Robust adaptive beamforming based on interference covariance matrix reconstruction and steering vector estimation. IEEE Trans. Signal Process. **60**(7), 3881–3884 (2012)

3. Vorobyov, S.A., Gershman, A.B., Luo, Z.Q.: Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem. IEEE Trans. Signal Process. **51**(2), 313–324 (2003)
4. Khabbazibasmenj, A., Vorobyov, S.A., Hassanien, A.: Robust adaptive beamforming based on steering vector estimation with as little as possible prior information. IEEE Trans. Signal Process. **60**(6), 2974–2987 (2012)
5. Chen, F.F., Shen, F., Song, J.Y.: Robust adaptive beamforming using low-complexity correlation coefficient calculation algorithms. IEEE Electron. Lett. **51**(6), 443–445 (2015)

Integration of Reed–Solomon Codes to Bundle Streaming Service (BSS) for Deep Space Communications

Qian Qu, Xiaoyu Zhu, Kanglian Zhao(✉), and Wenfeng Li

School of Electronic Science and Engineering, Nanjing University, 22 Hankou Rd., Nanjing 210093, Jiangsu, People's Republic of China
mf1523028@smail.nju.edu.cn, mf1723072@smail.nju.edu.cn, zhaokanglian@nju.edu.cn, leewf_cn@hotmail.com

Abstract. A novel data streaming scheme, simply named RS-BSS, is proposed based on Bundle Streaming Service (BSS) framework to perform high quality data streaming service in deep space. By integrating the Reed–Solomon (RS) codes into the BSS, the RS-BSS scheme can get more in-order data streaming within the initial transmission, which improves the performance of the "best-effort" service in BSS. Meanwhile, the "reliable" service in BSS can also be improved since the recovery capabilities provided by RS codes would reduce the retransmission times. By implementing the RS codes into the transport layer of BSS, the proposed scheme has the advantages of strong error recovery capabilities and better real-time performance. An emulated data streaming communication experiment between two nodes is conducted to evaluate the performance of the proposed scheme in deep space environment. The evaluation results show that the proposed scheme has less stream delivery time (SDT) and more stable end user's display efficiency (EDE) against the severe transmission conditions. This means the RS-BSS scheme improves the performance of both "best-effort" and "reliable" services in BSS.

Keywords: Reed–Solomon codes · Bundle Streaming Service (BSS) · Delay/disruption-tolerant network · Deep space

1 Introduction

As an increasing number of manned space program will be undertaken in the future, multimedia services in deep space such as real-time videos are becoming more and more important. However, due to the harsh transmission conditions including high loss of packets, frequent link disruptions, long propagation delays, and channel-rate asymmetry, it is difficult to perform high-quality data streaming service in deep space.

To face this challenge, several studies have been conducted to provide data streaming transmission based on delay/disruption-tolerant networking (DTN)

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 207–215, 2019.
https://doi.org/10.1007/978-981-13-6264-4_25

[1] framework, albeit not extensively. In [2], by embedding redundant "summary frames" to the original stream, T. Liu and S. Nelakuditi constructed a disruption-tolerant video sequence so that even in the disruption cases, the network may still delivery video content to clients. And, in [3], G. Karlsson *et al.* proposed a broadcasting system providing unreliable content distribution to a random block of receivers. Among these studies, Bundle Streaming Service (BSS) framework was proposed in 2013 [4] implementing a forwarding strategy including "best-effort" service and "reliable" service, which balances the trade off between efficiency in initial transmission and in retransmission, and also assures the in-order delivery of bundles. Meanwhile, the retransmission mechanism in BSS guarantees the integrity of the data at the receiver. These features make BSS a promising framework in deep space data streaming transmission. However, due to the severe transmission conditions in deep space, BSS may not perform well on real-time data delivery service in initial transmission due to the high packet loss rate. And the long propagation delay will make the retransmission expensive.

In this paper, a novel data streaming scheme is proposed to improve the real-time data delivery performance and reduce the retransmission times. The proposed scheme is implemented by integrating the Reed–Solomon (RS) codes into the BSS to restore the missing data at the receiver within the initial transmission.

The remaining sections of this paper are organized as follows. Section 2 briefly overviews the key principles of the BSS. Then, in Sect. 3, the design and implementation of RS-BSS are explicitly explained. In Sect. 4, a data streaming transmission experiment between two nodes is conducted to measure the performance of the RS-BSS scheme in deep space environment. The conclusions are drawn in Sect. 5.

2 Overview of BSS

Typical streaming data transmission in deep space is harmed by severe transmission conditions as mentioned before. To get over these difficulties, BSS is designed to approach the goal of unimpeded real-time streaming of all data delivered in transmission order, together with comprehensive replay and review of all the data in the stream [4]. The main features of BSS comprise the following three parts.

2.1 Forwarding Strategy

The architecture of BSS is shown in Fig. 1(a). BSS conducts the streaming functions at the Bundle layer and provides a bunch of transport services below the BSS forwarder including "best-effort" service and "reliable" service. Note that DTN implementations bind predetermined transport layer protocols to specific endpoints. Unlike that, BSS proposes a superb forwarding strategy, that is, to deliver most of the data with minimal delay and initial transmission overhead

through the "best-effort" service such as UDP and "green" LTP, while to assure the integrity of the data by retransmission through the "reliable" service such as TCP and "red" LTP. Besides, the flow of real-time data transmission neither waits for the lost bundles to be retransmitted nor displays the out-of-order bundles received, because in streaming data service such as video services, a few of data loss can be accepted. Rather, waiting for the retransmission would interrupt the continuity of the broadcast and degrade the viewing experience.

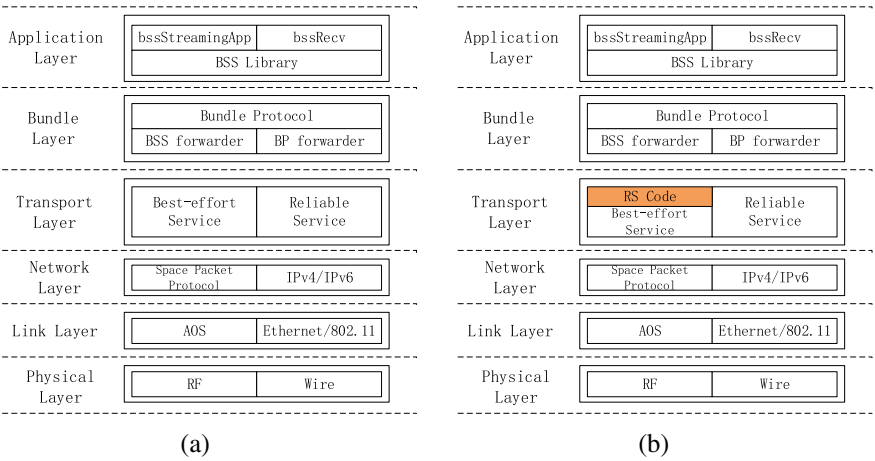


Fig. 1. The structure of (a) BSS, (b) RS-BSS

2.2 In-order Delivery

Unfortunately, DTN protocols, such as BP, do not perform any functionality for in-order transmission of data which is a critical point in data streaming propagation. In BSS, to handle this, the forwarder records the creation time of each bundle. At every hop along the end-to-end path, each bundle whose creation time is greater than that of all the other bundles seen on this stream is forwarded via the 'best-effort' service. Otherwise, it would be forwarded via the 'reliable' service [4]. Thus, only the in-order data will appear in the queue for real-time display.

2.3 Replay and Reliability

BSS guarantees the integrity of data at receiver by its retransmission mechanism. Each time it receives a bundle, no matter through "best-effort" service or through "reliable" service, it injects the data into the BSS database according to the order of its creation time for replay if necessary.

3 Design and Implementation of RS-BSS

The high loss rate of data in deep space, especially the burst errors, may deteriorate the quality of real-time streaming data delivery and even interrupt the real-time display at the receiver. Although the reliability of data transmission can be guaranteed by retransmission mechanism, it is an expensive approach considering the long propagation time in deep space.

To tackle this problem, a novel RS-BSS scheme is proposed to improve the real-time display quality and the transmission efficiency by integrating the RS codes into the "best-effort" service of BSS, as shown in Fig. 1(b). Note that the proposed scheme does not implement the code into the "reliable" service of the BSS such as TCP or "red" LTP, since the "reliable" service only performs to retransmit the lost data according to the forwarding strategy mentioned in Sect. 2. That means the "reliable" service exerts no influence on the real-time data transmission performance. And the innate ARQ mechanism in BSS would impose the retransmission automatically to ensure the integrity of the data. RS Code is a classical erasure code [5] widely applied in digital communications and data storage, which is also the Consultative Committee for Space Data Systems (CCSDS) space standard channel coding [1]. A RS code $RS(n, k)$, where n is the code length and k is the information length, can correct up to $(n-k)/2$ random errors and check up to $(n-k)$ errors [5].

The RS codes are performed into the transport layer for the following reasons: (i) it has stronger error correction capabilities allowing for block-sized correction abilities; (ii) it does not need to tackle the out-of-order bundles at the Bundle layer thanks to the forwarding strategy of BSS, which will contribute to the conciseness of the implementation; (iii) the decoding time could be ignored compared to the long propagation time in deep space.

Figure 2 shows the encoding and decoding flow of our scheme:

- Prior to the Reed–Solomon encoding process, the convergence layer first converges several bundles to a Block;
- Each Block is divided into k sub-Blocks. Note that the last sub-Block is often smaller than the others, so padding skills are used to expand it to the same size;
- These k sub-Blocks are sent to Reed–Solomon encoding procedure in order to generate the redundancy packets;
- Each sub-Block and the redundancy packet should be prepended a packet header before delivering to UDP. The information contained in a header comprises the index of the Block and the index of the packet (i.e., sub-Block and redundancy packet) which will help to perform the decoding procedure correctly at the receiver;
- At the receiving side, the receiver uses the coding information in the header to determine whether they belong to the same Block and then performs the decode process;
- If lost more than $(n-k)$ sub-Blocks belonging to same Block, the receiver cannot decode or check up the lost data, and thus informs the sender to retransmit the whole Bundle;

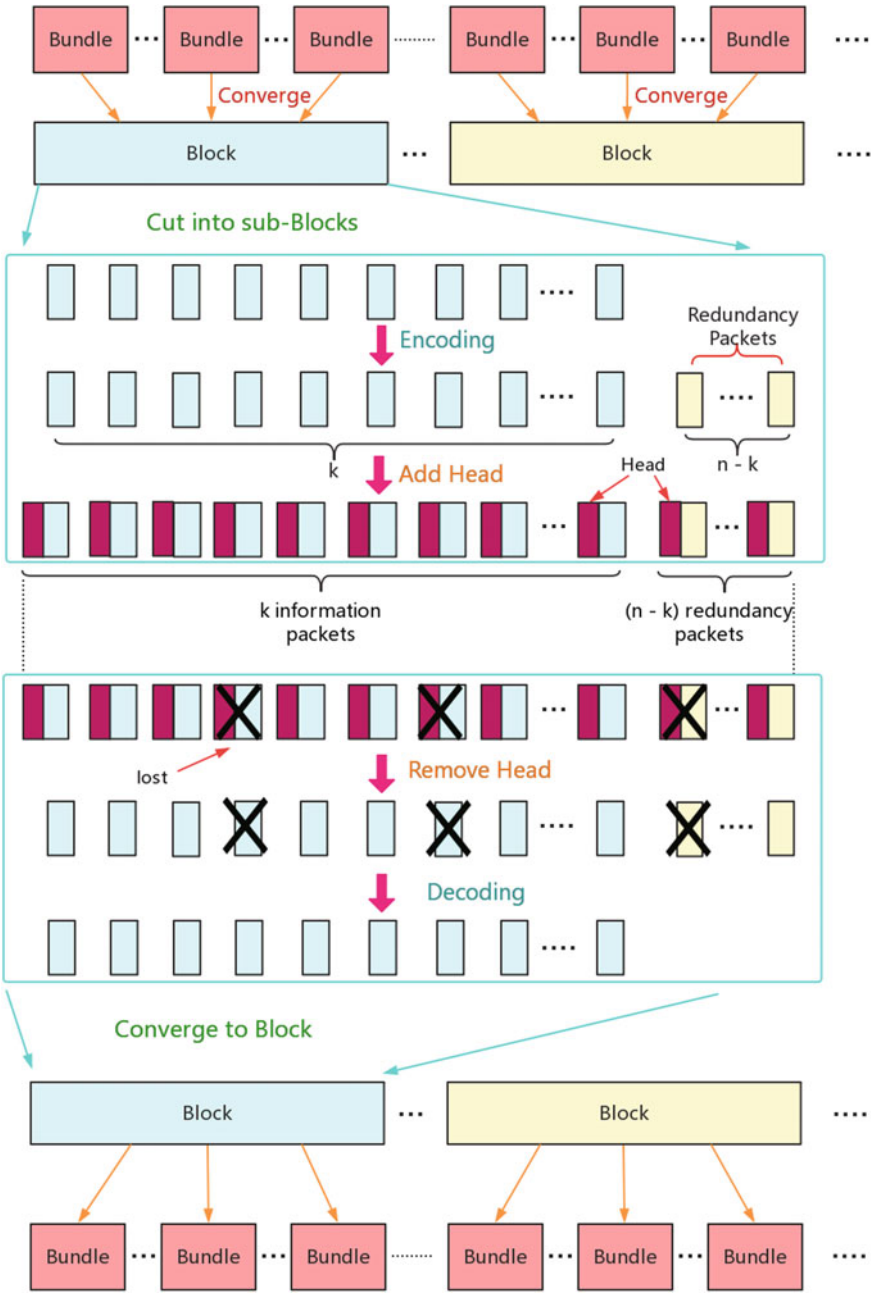


Fig. 2. The flow of Reed–Solomon Code implementation in transport layer

– If the lost sub-Blocks are more than $(n-k)/2$ and smaller than $(n-k)$, the receiver will inform the sender to retransmit specific data rather than resend the whole Bundle according to the original BSS;
– Otherwise, the decoder will recover the lost sub-Blocks and present them to the upper layer for convergence and real-time display (Table 1).

Table 1. Experimental factors and configuration

Experimental factors	Settings/values
ION version	3.6.0b
Operating System	Linux, kernel 3.19.0
Data bundle size(bytes)	11844
Highest decoding rate(Mbps)	2800
Code rate	0.6
BER	10^{-7}, 5×10^{-7}, 10^{-6}, 5×10^{-6}
One-way link delay(s)	$1 \sim 40$
Stream rate(Mbps)	3
Amount of stream data	20s stream data
Sample size	10 repetitive runs

4 Evaluation Results

To evaluate the performance of proposed RS-BSS scheme over the original BSS, the following metrics [4] were exploited for evaluation: (i) stream delivery time (SDT), which is the duration for all the data to be successfully delivered to the receiver; and (ii) end user's display efficiency (EDE), which is the ratio of displayed video frames to all transmitted frames within the initial transmission.

To implement an emulated deep space streaming data delivery environment, three PCs were deployed to set up the experiments, where two of them were equipped with BSS, representing the source node and the end node respectively, and the other was used as a channel emulator to control the changes of channel error rates and propagation delays in the experiments. The BSS implementation was provided by the Interplanetary Overlay Network (ION) v3.6.0b [7]. Two simple streaming applications provided by ION were also modified to emulate the behavior of real streaming applications where "bssStreamingApp" simulated HD videos (i.e., H.264 1280x720@30.0 fps) by producing a stream of 30 frames per second at a constant bit rate of 3 Mbps. Meanwhile, the "bssRecv" presented two primary functions with the help of the BSS library: (i) the real-time video services by displaying any in-order bundles once received, and (ii) the replay services by saving all bundles to a specially designed database and sorting them

by transmission order. In the experiment, each bundle's payload was a multiple of 1316 bytes (e.g., 11844 bytes in our experiment) in order to simulate the behavior of the packetized elementary stream mechanism [8] widely used in popular media streaming applications such as VLC [9]. Wireshark was used to detect packages sent and received from both ends, thus helping to determine the delivery time of our stream data. Each test was repeated 10 times to preclude hidden factors.

In the first phase of the evaluation, the SDT time metric was employed to measure the network performance. As the results illustrated in Fig. 3(a) and (b), the RS-BSS scheme outperforms BSS in most cases, managing to deliver all the bundles within a shorter time. In particular, it is observed in Fig. 3(a) that in both moderate and severe channel conditions, RS-BSS offers far faster data delivery than the original BSS does. Furthermore, the performance gap is up to four times between two schemes when the propagation delay increases, as shown in Fig. 3(b). It can be seen that when the propagation delay is 1 s, the RS-BSS scheme can deliver the full data within 70 s even in severe channel conditions, while BSS takes as much as 3 times up to 202 s in total. And when the propagation delay is 30 s, BSS also takes more time to deliver the data, up to 7 min in severe channel conditions.

In the multimedia performance evaluation, it is apparent that RS-BSS achieves better performances of display by getting more media bundles in the receiver within the initial transmission. As described in Fig. 4(a) and (b), compared with BSS, our proposed RS-BSS scheme can get relatively stable end user's display efficiency (EDE), even under the harsh transmission conditions such as longer propagation delay and higher bit error rate.

The results above seem reasonable, since the coding strategy would significantly reduce the number of retransmission bundles, which will contribute exponentially to the SDT in deep space. That means the RS codes would benefit to the "reliable" service in BSS. What is more, the resilience of the data in "best-effort" service would benefit the initial multimedia transmission as well as the robustness against the harsh channel environment.

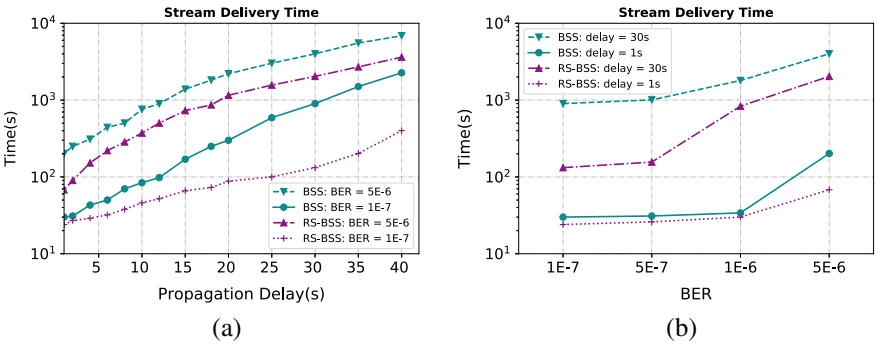


Fig. 3. Comparison between BSS and RS-BSS based on stream delivery time metric respect to (a) propagation delay (b) BER

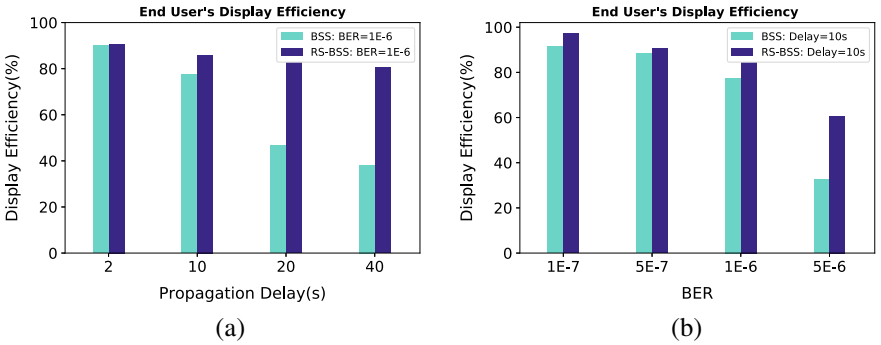


Fig. 4. Comparision between BSS and RS-BSS based on end user's display efficiency metric respect to (a) propagation delay (b) BER

5 Conclusion

In order to improve the real-time transmission quality and the data streaming transmission efficiency, we presented a novel scheme by integrating the RS codes into the original BSS. The experiment results show that the proposed RS-BSS scheme achieves better performance in terms of the stream delivery time (SDT) and the end user's display efficiency (EDE). Due to the recovery capabilities RS codes furnished, the proposed scheme spreads out higher real-time data transmission efficiency which will improve the performance of "best-effort" service. Meanwhile, the scheme can also perform as an important supplement to retransmission mechanism contributing to the performance of "reliable" service.

In the future, more transport protocols such as LTP and more FEC codes such as LDPC would be considered in our scheme, and more experiments will be implemented to evaluate the performance of the proposed scheme.

References

1. The Space Internet working Strategy Group (SISG), Recommendations on a strategy for space Internet working, IOAG.T.RC.002.V1, Report of the Interagency Operations Advisory Group, NASA Headquarters, Washington, DC 20546-0001, USA, 1 Aug 2010 (2010)
2. Liu, T., Nelakuditi, S.: Disruption-tolerant content-aware video streaming. In: MULTIMEDIA 2004: Proceedings of the 12th Annual ACM International Conference on Multimedia, New York, NY, USA, pp. 420–423 (2004)
3. Karlsson, G., Lenders, V., May, M.: Delay-tolerant broadcasting. In: Proceedings of 2006 SIGCOMM Workshop on Challenged Networks, Pisa, Italy, pp. 197–204 (2006)
4. Sotirios-Angelos, L., Scott, C.B., Vassilis, T.: Bundle streaming service: design, implementation and performance evaluation. Trans. Emerging Tel. Tech. **26**, 905–917 (2015)

5. Reed, I.S., Solomon, G.: Polynomial codes over certain finite fields. J. Soc. Ind. Appl. Math. **8**(2), 300–304 (1960)
6. Cola, T.: A protocol design for incorporating erasure codes within CCSDS: the case of DTN protocol architecture. In: 5th Advanced Satellite Multimedia Systems Conference and the 11th Signal Processing for Space Communications Workshop, pp. 68–73 (2010)
7. Interplanetary Overlay Network (ION). Jet Propulsion Laboratory. California Institute of Technology, CA. http://ion-dtn.sourceforge.net
8. Cisco Corporation. Fundamental of Digital Video Cisco. http://www.cisco.com/en/US/docs/solutions/Enterprise/Video/pktvideoaag.pdf
9. VideoLan Media Player. http://www.videolan.org/vlc/index.html

Secure Noise-and-Forward Relaying with Polar Codes

Bin Duo(✉), Peng Zhang, Yanmei Hu, Yong Fang, Yong Jia, and Junlin Xiang

Chengdu University of Technology, Chengdu 610059, China
duo_bin@163.com

Abstract. A realizable polar coding scheme for secure communications in the four-node relay-eavesdropper channel (FNREC) is proposed in this paper. Based on the noise-and-forwarding (NF) cooperative strategy, the relay only serves as a disturber to the eavesdropper, which makes the eavesdropper channel to be degraded with respect to the legitimate channel. The corresponding code construction is designed, and the encoding and decoding procedure are provided. It is shown that our scheme can achieve the maximum perfect secrecy rate and provide both reliability and security over the FNREC with low encoding and decoding complexity.

Keywords: Physical-layer security · Cooperation · Noise-and-forwarding (NF) · Perfect secrecy rate · Polar codes

1 Introduction

From an information-theoretic point of view, physical-layer security has been regarded as one of the fundamental problems in the wireless networks and has attracted much attention in recent years. The notion of physical-layer security was first introduced by Shannon [1]. Then, Wyner [2] studied the wiretap channel and stated that secure communications can be achieved by channel coding at the physical layer.

More recently, considerable research on the secure communications has been focused on multiple antenna systems [3–6]. However, due to the cost and size limitations, using multiple antennas at each node may not be practically feasible. Cooperative communications as an effective technique make each node in wireless networks to be equipped with only single antenna, so that they can also enjoy the benefits of multiple antenna systems [7]. Relay networks are important components of cooperative communications in which some nodes act as relays to assist the transmission between two nodes of networks.

As a natural extension, relays are used to improve the secrecy level of communications in relay networks in the existence of eavesdroppers in order to solve physical-layer security issues. The model of four-node relay-eavesdropper channel (FNREC) was first introduced by [8]. In particular, the corresponding perfect secrecy rate based on the proposed noise-and-forwarding (NF) cooperation strategy was derived. The idea of NF is that the relay cannot hear any information about the confidential message from the source but only transmits random codewords to jam the eavesdropper. However,

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 216–223, 2019.
https://doi.org/10.1007/978-981-13-6264-4_26

this work relies on non-realizable random-coding technique to prove the results. These results show the existence of some channel coding schemes that achieve the perfect secrecy rate, but they do not show how to design specific channel codes with a low encoding and decoding complexity.

Very recently, polar codes as a practical channel coding scheme via the method of channel polarization have been introduced in [9]. The construction method of polar codes makes them applicable to secure communication scenarios. Polar codes for multi-user channels have been studied broadly in [10–15]. In particular, [15] proposed a secure transmission scheme for the degraded FNREC by using polar codes and proved that the scheme achieves both reliability and security under the decode-and-forward (DF) strategy. However, if the source-to-relay (SR) channel is blocked, the DF strategy may not be applicable. In this case, we let the relay send independent noise signals to confuse the eavesdropper based on the NF strategy. It is shown that a positive secrecy rate may be achieved even when the source-to-eavesdropper (SE) channel is better than the source-to-destination (SD) channel [15].

To the best of knowledge, polar coding for secure transmission using NF strategy has not been addressed in the literature. In this paper, we design polar coding scheme and present the encoding and decoding procedure. It is shown that the maximum perfect secrecy rate can be achieved by employing the NF strategy over the FNREC while keeping a remarkably low encoding and decoding complexity. Moreover, from the analysis and the simulation results, a conclusion can be arrived at that the proposed approach ensures both the reliability and security.

2 System Model

We first consider the secure communication scenario in the FNREC with binary-input alphabet. For the NF strategy, the source (S) would like to deliver a confidential message M to the legitimate destination (D) with a transmission rate as high as possible while keeping M secret from the illegitimate eavesdropper (E). The relay plays the role of a "confuser" according to the NF protocol. More specifically, S encodes the message M into a length-N codeword X_S, and broadcasts it into the FNREC. The coding rate is denoted by R_S. At the same time, R transmits random codewords to confuse E. D receives $Y_{D,1}$ and $Y_{D,2}$ from S and R, respectively, and recovers the estimated message $\hat{M}$. E observes $Y_{E,1}$ and $Y_{E,2}$ from S and R, respectively. We assume that D and E receive signals from orthogonal channels.

For this FNREC, the reliability condition is measured by the error probability

$$\lim_{N\to\infty} P_e(N) = \lim_{N\to\infty} \Pr(\hat{M} \neq M) = 0 \tag{1}$$

The secrecy condition can be evaluated by the equivocation rate at E, which is given by

$$R_e = \frac{1}{N} H(M|Y_{E,1}, Y_{E,2}) = \frac{1}{N} H(M) = R_S \tag{2}$$

The equivocation rate R_e describes the remaining uncertainty of E on M after observing $Y_{E,1}$, $Y_{E,2}$. If R_S is achievable for $N \to \infty$, i.e. $R_S = R_e$, we say that R_S is the perfect secrecy rate.

Let W_{SD}, W_{SE}, W_{RD}, and W_{RE} be the SD channel, SE channel, relay-to-destination (RD) channel and relay-to-eavesdropper (RE) channel in the FNREC, respectively. We consider the case that W_{SD} is much nosier than W_{SE} (i.e. $W_{SD} \preceq W_{SE}$) and the channel W_{RD} condition is better than that of W_{RE} (i.e. $W_{RE} \preceq W_{RD}$). Note that if $W_{RD} \preceq W_{RE}$, it is hard to achieve the positive secrecy rate. In the NF strategy, the relay dose not listen any information from the source but transmits independent random bits in order to confuse the eavesdropper. Theorem 1 below provides a closed-from expression for the perfect secrecy rate of the NF strategy in the FNREC.

Theorem 1: For the FNREC using the NF strategy, if $W_{SD} \preceq W_{SE}$ and $W_{RE} \preceq W_{RD}$, the following maximum perfect secrecy rate is achievable

$$R_S^{NF} = I(W_{SD}) + I(W_{RD}) - (I(W_{SE}) + I(W_{RE})) \tag{3}$$

where $I(W_{ij})$ is the corresponding mutual information of the channel W_{ij} for $i \in \{\mathrm{S,R}\}$, $j \in \{\mathrm{R,D}\}$.

The proof can be straightforwardly derived from the proof of Theorem 1 in [15].

3 Secure Polar Coding Scheme

Though the NF strategy is shown to achieve secrecy capacity by using random-coding technique [8], how to design practical secrecy capacity-achieving codes with realizable encoding/decoding complexity for the FNREC is an elusive task. In this section, the remarkable characteristics of polar codes are used to address this problem.

In our model, since $W_{SD} \preceq W_{SE}$, which means that the legitimate channel has a lower signal-to-noise ratio (SNR) than the eavesdropper channel, the confidential message cannot be kept secure from E. To solve this issue, the relay sends independent codeswords in order to degrade the eavesdropper channel and assist in confusing E.

Hence, the signals received by D and E are given by

$$Y_D = h_{SD}\sqrt{P_S}X_S + h_{RD}\sqrt{P_R}X_R + N_D \tag{4}$$

$$Y_E = h_{SE}\sqrt{P_S}X_S + h_{RE}\sqrt{P_R}X_R + N_E \tag{5}$$

where N_D and N_E are additive white Gaussian noises (AWGN) of the legitimate channel and the eavesdropper channel respectively, h_{ij} ($i \in \{\mathrm{S,R}\}$, $j \in \{\mathrm{R,D}\}$) is the channel gain for the corresponding channel, and P_S and P_R are the transmit power of S and R, respectively. It is clear from (4) and (5) that by adjusting P_S and P_R, we can guarantee that the eavesdropper channel is degraded with respect to the legitimate channel even if the original eavesdropper channel is better, because X_R is known to the

destination, which can be cancelled at D. In this case, the eavesdropper channel is degraded by the independent codewords with high probability while the legitimate channel remains unaffected.

Moreover, we assume that the source has perfect channel state information of all the channels, so all the channels are equivalent to binary-input AWGN channel. By using the NF protocol, W_{SD}, W_{SE} and W_{RD}, W_{RE} are equivalent to two different binary-input AWGN channels, i.e. W_D and W_E, which satisfies $W_D \preceq W_E$.

Thus, the design of polar codes for the FNREC fading channels with the NF protocol is similar to the polar coding design for binary-input AWGN wiretap channels. We show that in this case our polar coding scheme achieves the secrecy rate with binary-input constraint.

3.1 Construction of Polar Codes

According to Lemma 2 in [15], we know that if $W_D \preceq W_E$, the corresponding information sets denoted by $\mathcal{A}_D$ and $\mathcal{A}_E$, respectively, satisfy $\mathcal{A}_E \subseteq \mathcal{A}_D$. Therefore, three independent subsets of $[N]$ at the source can be defined as illustrated in Fig. 1: $\mathcal{A}_D \backslash \mathcal{A}_E$ means the information set for W_{D} and the frozen set for W_{E}; $\mathcal{A}_E$ is information set for both of the channels W_{D} and W_{E}; $\mathcal{F} = [N] \backslash \mathcal{A}_D$ represents the frozen set for both of the channels W_{D} and W_{E}.

Note that in the conventional wiretap channel, the source not only sends its own confidential message, but also sends extra random bits to jam the eavesdropper. In our scheme, the role of jammer can be replaced by the relay sending independent codewords in order to confuse the eavesdropper. That is the reason why the scheme in Fig. 1a can be transformed to the equivalent binary-input AWGN wiretap channel as shown in Fig. 1b.

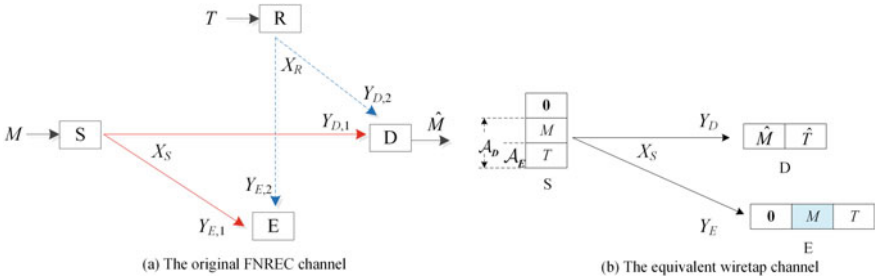


Fig. 1. The construction of polar codes for the FNREC by using the NF strategy

3.2 Encoding and Decoding

Encoding: Since $W_D \preceq W_E$, the source message is divided into two independent message, i.e. M and T, respectively. To ensure security, the source generates the

confidential message M uniformly from $\{0,1\}^{|\mathcal{A}_D \backslash \mathcal{A}_E|}$ and generates uniformly random bits T from $\{0,1\}^{|\mathcal{A}_E|}$ to confuse the eavesdropper, which satisfies

$$\lim_{N\to\infty} \frac{|\mathcal{A}_E|}{N} = I(W_{SE}) + I(W_{RE}) \tag{6}$$

Then the channel input is constructed as

$$X_S = MG_{\mathcal{A}_D \backslash \mathcal{A}_E} \oplus TG_{\mathcal{A}_E} \tag{7}$$

where $G_{\mathcal{A}_D \backslash \mathcal{A}_E}$ and $G_{\mathcal{A}_E}$ are the corresponding generator matrix for encoding M and T, respectively. As a result, the polar coding rate at the source is given by

$$\begin{aligned} \lim_{N\to\infty} R_S &= \lim_{N\to\infty} \frac{|\mathcal{A}_D \backslash \mathcal{A}_E|}{N} = \lim_{N\to\infty} \left(\frac{|\mathcal{A}_D|}{N} - \frac{|\mathcal{A}_E|}{N}\right) \\ &= I(W_{SD}) + I(W_{RD}) - (I(W_{SE}) + I(W_{RE})) = R_S^{NF} \end{aligned} \tag{8}$$

This means the maximum secrecy rate can be achieved by our coding scheme.

Decoding: According to the observation Y_D, D first recovers both $\hat{M}$ and $\hat{T}$, the estimates of M and T, together by using the successive cancellation decoding algorithm [9]. Then, it extracts $\hat{M}$ according to the code construction. It is straightforward from the capacity-achieving property of polar codes that the error probability the destination decodes can be upper bounded by $O(2^{-N^\beta})$ by choosing $\beta < 1/2$ and sufficiently large N. Thus, the proposed coding approach satisfies the reliability condition (1). The complexities of polar encoding and decoding are both $O(N\log N)$ [9].

It has been proved in [11] that polar codes for the general wiretap channel satisfies the security constraint of (2). Therefore, we skip the proof here.

From the information-theoretic point of view, the proposed approach is asymptotically optimal. When the block length of polar codes N goes to infinity, the proposed approach can ensure both of the reliability and the security of communications and can also achieve the secrecy rate of the FNREC with low encoding and decoding complexity.

4 Simulation Results

From the analysis above, D can decode the confidential message with any small error probability using the proposed approach when the block length goes to infinity. However, it is also of interest to validate the bit error rate (BER) performance of polar codes under the SC decoding in the finite regime.

We use the simulation specification that is listed in Table 1 in the following simulation.

Table 1. Simulation parameters

Parameters	Specification
M	2000 bits
Block lengths	$N = 2^8, 2^9, 2^{10}, 2^{11}, 2^{12}, 2^{13}, 2^{14}$
$I(W_{SD})$	0.1
$I(W_{SE})$	0.2
$I(W_{RD})$	0.5
$I(W_{RE})$	0.2

Figure 2 shows the BER performance of the legitimate channel and the eavesdropper channel for different block lengths. As expected, with the increase of the code length, the BER of the legitimate destination decreases apparently when decoding the received confidential message. Meanwhile, the BER at the eavesdropper remains almost stable at around 0.5. This is because the eavesdropper cannot obtain enough useful information to decode the confidential message based on the proposed scheme. The poor performance for short and moderate codes is a common phenomenon of polar codes for finite block lengths due to the insufficiency of channel polarization compared to full polarization when the block length goes to infinity. It is expected, the proposed approach will achieve the maximum perfect secrecy rate with any small bit error rate when the block length is growing larger enough.

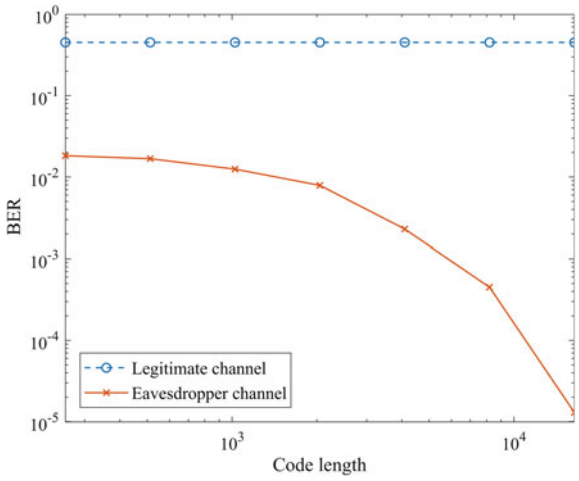


Fig. 2. BER performance of the proposed scheme

5 Conclusion

In this study, polar codes have been shown to be suitable for noise-and-forward cooperation strategy in the FNREC. In particular, the method of the code construction has been proposed and the encoding/decoding processes have been developed for this channel. It is shown than our scheme achieves the maximum perfect secrecy rate with low encoding and decoding complexity. Numerical results and the analysis have validated the proposed approach that can ensure both the reliability and security of communications.

Acknowledgement. This research work is supported by the National Natural Science Foundation of China (61101125), Scientific Research Foundation of the Science & Technology Department of Sichuan Province, China (18ZDYF3446), Scientific Research Foundation of the Education Department of Sichuan Province, China (18ZB0066), Scientific Research Foundation of the Science and Technology Department of Chengdu, China (2017-RK00-00363-ZF).

References

1. Shannon, C.: Communication theory of secrecy systems. Bell Syst. Tech. J. **28**, 656–715 (1949)
2. Wyner, A.D.: The wire-tap channel. Bell Syst. Tech. J. **54**(8), 1355–1387 (1975)
3. Khisti, A., Wornell, G.W.: Secure transmission with multiple antennas-Part II: The MIMOME wiretap channel. IEEE Trans. Inf. Theory **56**(11), 5515–5532 (2010)
4. Khisti, A., Wornell, G.W.: Secure transmission with multiple antennas I: The MISOME wiretap channel. IEEE Trans. Inf. Theory **56**(7), 3088–3104 (2010)
5. Chen, J., Chen, X.M., Liu, T., Lei, L.: Toward green and secure communications over massive MIMO relay networks: Joint source and relay power allocation. IEEE Access **5**, 869–880 (2017)
6. Chen, X.M., Kwan Ng, D.W., Gerstacker, W.H., Chen, H.H.: A survey on multiple-antenna techniques for physical layer security. IEEE Commun. Surv. Tut. **19**(2), 1027–1053 (2017)
7. Laneman, J.N., Tse, D.N.C., Wornell, G.W.: Cooperative diversity in wireless networks: efficient protocols and outage behaviour. IEEE Trans. Inf. Theory **50**(12), 3062–3080 (2004)
8. Lai, L., El Gamal, H.: The relay-eavesdropper channel: cooperation for secrecy. IEEE Trans. Inf. Theory **54**(9), 4005–4019 (2008)
9. Arikan, E.: Channel polarization: a method for constructing capacity achieving codes for symmetric binary-input memoryless channels. IEEE Trans. Inf. Theory **55**(7), 3051–3073 (2009)
10. Chou, R.A., Bloch, M.R.: Polar coding for the broadcast channel with confidential messages: a random binning analogy. IEEE Trans. Inf. Theory **62**(5), 2410–2429 (2016)
11. Wei, Y.P., Ulukus, S.: Polar coding for the general wiretap channel with extensions to multiuser scenarios. IEEE J. Sel. Areas Commun. **34**(2), 278–291 (2016)
12. Zhao, Y.Z., Zou, X.C., Lu, Z.J., et al.: Chaotic encrypted polar coding scheme for general wiretap channel. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **25**(12), 3331–3340 (2017)
13. Zhang, Y.X., Yang, Z., Liu, A.J.: Secure transmission over the wiretap channel using polar codes and artificial noise. IET Commun. **11**(3), 377–384 (2017)

14. Liu, L., Yan, Y.F., Ling, C.: Achieving secrecy capacity of the Gaussian wiretap channel with polar lattices. IEEE Trans. Inf. Theory **64**(3), 1647–1665 (2018)
15. Duo, B., Wang, P., Li, Y.H., et al.: Secure transmission for relay-eavesdropper channels using polar coding. IEEE ICC, pp. 2197–2202 (2014)

Design of Rate-Compatible Punctured Algorithm for Irregular LDPC Codes

Honglin Zhao(✉), Jianting Huang(✉), and Haiyue Zhang(✉)

Harbin Institute of Technology, Xidazhijie, 92, Harbin 150006, China
hlzhao@hit.edu.cn, {jianting_huang, zhyhit}@foxmail.com

Abstract. The excellent error correction performance of Low-Density Parity Check code has made it widely used in many modern communication systems. In time-varying fading channels, multiple sets of codecs with different code rates are often used to adapt to different states of the channel. And rate-compatible coding is an effective implementation of the problem. This paper proposed an improved centralized puncturing algorithm applied to irregular LDPC. And the proposed algorithm is verified by comparing the performance of punctured LDPC codes with CCSDS standard and Ma Fuli's algorithm.

Keywords: LDPC · Rate-compatible · Puncturing

1 Introduction

In time-varying channels, channel coding must not only have strong error correction capability, but also dynamically adjust the code rate of the transmitted code according to the current state of the channel. This has a higher requirement for the structural design of the channel coding. In 1988 Hagenauer first proposed Rate-Compatible Punctured Convolutional Code [1]. Its greatest advantage is that it can use a pair of codecs to perform coding and decoding, reducing hardware resource consumption and system implementation complexity. This provides new ideas for the development of later error correction codes, such as Rate-Compatible Punctured Turbo [2] and Rate-Compatible LDPC [3].

The error correction performance of LDPC codes is able to approach the Shannon limit [4], which makes it be considered for a wide range of applications in practical engineering. The LDPC code was first proposed by Gallager [5] in 1962. In 1981, Tanner [6] introduced the bipartite graph into the analysis of LDPC codes. For the implementation of rate-compatible code in 2004, J. Ha et al. proposed the concept of recovery tree for the first time and adopted this concept to select the punctured bits[7, 8]. In 2011, Ma Fuli proposed a centralized puncturing algorithm based on the recovery tree which is applicable to regular LDPC codes [9].

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 224–231, 2019.
https://doi.org/10.1007/978-981-13-6264-4_27

This work suggests a centralized puncturing algorithm for an irregular LDPC with a code length of 2560 and a code rate of 0.4 and verifies the performance advantages through MATLAB simulation.

2 Basic Concept for Puncturing Algorithm

In this section, a brief introduction of the puncturing algorithm is given, including the definition and facts. And the concept of Recovery Tree is explained concisely, which is the most basic concept of the algorithm in this paper.

2.1 Recovery Tree

n LDPC code is a linear block error correction code that can be defined by a sparse parity check matrix $H$of dimensions $m \times n$ or a bipartite graph. The n columns of H correspond to variable nodes $(v_1, v_2, \ldots, v_n)$ respectively in the bipartite graph, and the m rows correspond to check nodes $(c_1, c_2, \ldots, c_m)\cdot v_j(1 \leq j \leq n)$ and $c_i(1 \leq i \leq m)$ is connected in the Tanner graph if $H_{ij} \neq 0$. For an arbitrary node v we define N_v as a set of all nodes that can be reached from v by traversing only one edge. The nodes in N_v and the size of N_v are called the neighbors of v and the degree of v, respectively.

Suppose that some variable nodes are punctured. The decoder will set their log-likelihood ratios (LLRs) to 0 at the initialization. After iteration, all punctured variable nodes will be recovered by the information of the unpunctured nodes.

Through iteration, if a check node associated with the punctured nodes has only one neighbor punctured node, the check node will be referred to survived check node. If a check node associated with not only one punctured variable node, the check node will be called as dead check node, according to the check node cannot decide its message to the punctured node. However, as the decoding iteration proceeds, the punctured node associated with the "dead check node" will gradually recover, and it will become "live" when it associates only one puncturing node, i.e., the survived check node.

We classify the punctured variable nodes according to their recovery order in the decoding process. The punctured nodes that can obtain nonzero information in the first iteration process are called one-step-recoverable (1-SR). The punctured nodes recovered during the kth iterative process is called k-step-recoverable (k-SR). And 0-SR denotes unpunctured variable nodes. All variable nodes are grouped by k value, we put the unpunctured node in G_0, and put k-SR in G_k.

For a k-SR node v, we create a recovery tree originating from v in the following way. First, we link v with its guaranteed survived check node p and link p with all variable nodes in the set $N_p \backslash \{v\}$. Next, repeat the step on every new punctured variable node in the tree until every branch terminates with an unpunctured variable node. The resulting tree is called the recovery tree of v. Figure 1 is a 3-SR node recovery tree.

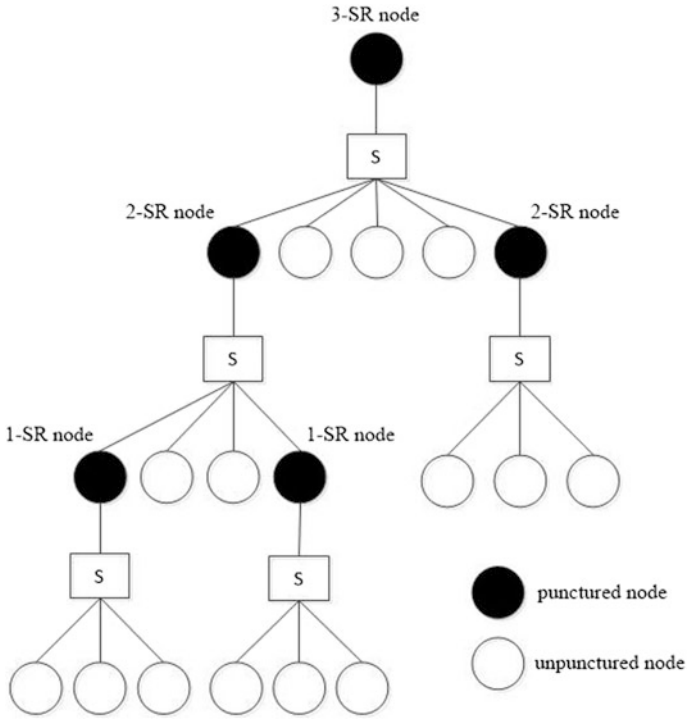


Fig. 1 Recovery tree of 3-SR node

2.2 Definition and Facts

In the recovery tree, the number of unpunctured nodes is very important. We define the number of unpunctured nodes in the recovery tree of v as $S(v)$. As the decoding iteration progresses, the variable node information is continuously updated. When the punctured node is recovered, we define its recovery error probability as follows.

For $v \in G_k$ and $k \geq 1$, $P_e(v)$ is the probability that v is recovered with a wrong message in the kth iteration, where the message is based only on information from the unpunctured nodes in the recovery tree of v.

For example, we transmit a punctured LDPC code on a BEC (Binary Erasure Channel), the probability of channel erasure is ζ, then for a variable node in G_k, its recovery error probability is expressed as

$$\psi(v,\zeta) = \begin{cases} (1-\zeta) & \text{for } v \in G_0 \\ \prod_{j=1}^{d_c-1} \psi(\gamma_j,\zeta) & \text{for } v \in G_k \text{ and } k > 0 \end{cases}$$

where G_0 and G_k are sets of unpunctured nodes and k-SR nodes, respectively, d_c is the degree of the survived check node of v, γ_j's are the neighbors of the survived check

node except for v, and $\gamma_j \in G_h$ for $0 \le h \le k-1$. Then the recovery error probability of a variable node can be simply expressed as $(1-\psi(v,\zeta))/2$.

From the above analysis, we come to the following conclusions:

Fact 1: On a BEC with an erasure probability of ζ, the recovery error probability $P_e(v)$ is $(1-\psi(v,\zeta))/2$, where $\psi(v,\zeta)=(1-\zeta)^{S(v)}$ for $v \in G_k$ and $k \ge 1$, $\psi(v,\zeta)$ is the same as the probability that the variable node v is recovered.

Fact 2: On a BSC (Binary Symmetric Channel) with an erasure probability of ζ, the recovery error probability $P_e(v)$ is $(1-\psi(v,2P_e))/2$, where $\psi(v,\zeta)=(1-\zeta)^{S(v)}$ for $v \in G_k$ and $k \ge 1$.

From the above facts, we can tell that a k-SR node with larger $S(v)$ has a larger recovery error probability. So we will look for a variable node with a smaller $S(v)$ as a new k-SR node. The recovered information of the first iteration recovery node is from the channel, which is relatively reliable. Then the recovery of the node during the second iteration, some of which the recovery information comes from the nodes recovered in the first iteration, so the reliability is not as reliable as the information recovered from the nodes in the first iteration.

3 Algorithm for Determining Puncturing Patterns

First of all, if R_0 represents the code rate of the mother code with code length of N, and P_m represents the number of punctured bits. Then the code length after punching is $N-P_m$, and the code rate is increased to $R_m = \frac{NR_0}{N-P_m}$. Therefore, the number of punctured bits can be determined according to the required code rate by the following equation:

$$P_m = \left\lfloor \frac{N(R_m - R_0)}{R_m} \right\rfloor$$

According to J. Ha et al., those variable nodes that are to be recovered by less iterations should be preferentially selected as punctured bits. In Ma Fuli's paper, we have seen a similar idea, that is, to make the number of 1-SRs as much as possible. These considerations are to recovery the punctured bits as soon as possible, and the effect of the punctured nodes on other nodes is as small as possible. The second consideration is to select a new punctured node among the neighbor nodes of the check node set with the least number of associated punctured nodes. The check node set is referred to P_{check}. And the set of variable nodes associated with check nodes in P_{check} is referred to P_{var}. This is in consideration of minimizing the impact of newly-selected punctured nodes on existing punctured nodes. This paper makes improvements based on the algorithm. That is, improve the strategy of selecting new punctured bit in P_{var} so that it can be applied to irregular LDPC codes.

Before describing the algorithm we define some basic symbols as follows. C represents the set of all check nodes, P represents the number of punctured nodes required to achieve the specified code rate, c_p represents the number of punctured nodes

associated with the check nodes p, P_{check} represents the set of check nodes associated with least punctured nodes, P_{var} represents a set of variable nodes associated with check nodes in P_{check}, v_n represents the number of check nodes in P_{check} associated with variable node v, *cweight* represents column weight of H, *INDEX* represents the column index set of punctured nodes, *Guard* represents a sentinel status in the algorithm, the specific role will be explained in detail later.

3.1 Proposed Centralized Puncturing Algorithm

The algorithm steps are as follows:

1. Initialization, $p = 0, INDEX = \varnothing, Guard = 0$
2. for($p = 0, p < P, p++$)
3. for $c \in C$, do
4. calculate the number of punctured nodes associated with each check node
5. end for
6. select the check node with the smallest c_p to create a set $P_{check} = \{c : c = \arg\min_{c' \in C} c'_p\}$
7. based on set P_{check} create a set of variable nodes, $P_{\text{var}} = \bigcup_{c \in P_{check}} N(c)$
8. for v $\in P_{\text{var}}$, do
9. calculate v_n, the number of check nodes in P_{check} associated with variable node
10. end for
11. if $Guard == 0$
12. $flag = 0$
13. if ($v_n == cweight(n)$) && (v_n should be as small as possible)
14. add this variable node to $INDEX$, $p++, flag = 1$
15. end if
16. if $flag == 0$
17. $Guard = 1$
18. end if
19. end if
20. if $Guard == 1$
21. finding variable nodes with as few v_n as possible and as much column weight as possible
22. add this variable node to $INDEX$, $p++$
23. end if
24. end for

3.2 Details of the Centralized Puncturing Algorithm

The input to the algorithm is a parity check matrix, and the output is a set of punctured node column indexes. The algorithm initially calculates the number of punctured nodes associated with each check node. Then group the check nodes associated with the least punctured node into a set P_{check}. When *Guard* is equal to zero, the variable node with the smallest v_n is selected as a punctured node in the set where v_n is equal to the column weight. This choice fits the rule that makes the association between the newly selected punctured node and the associated punctured node check equation as small as possible. When *Guard* is not equal to zero, the node with v_n as small as possible and the column weight as large as possible is punctured, in order to affect more recovery trees, so that $S(v)$ in each recovery tree is minimized.

4 Simulations

The (2048, 1024) LDPC code under the CCSDS(The Consultative Committee for Space Data Systems) is a kind of puncturing code, and its one frame codeword is obtained by deleting the last 512 bits of a 2560-bit code word. And its code rate is increased from 0.4 to 0.5. In this paper, the (2560, 1024) LDPC code under the CCSDS standard is selected as the mother code.

First, we implement mother LDPC code at rate 0.4, which is irregular and its block length is 2560 as mentioned before. Next, we puncture the mother code to obtain punctured LDPC codes at rate 0.5, 0.6, 0.7, and 0.8. The block lengths of punctured LDPC codes and the number of punctured bits for the rate are listed in Table 1. We have also simulated Ma Fuli's algorithm at the corresponding code rate. For comparison, we also simulated the performance of the punctured code at rate 0.5 under the CCSDS standard.

Table 1 Block length of punctured LDPC codes; The length in parentheses are the numbers of punctured symbols at the rates

Block lengths	Code rates				
	0.4	0.5	0.6	0.7	0.8
2560	2560 (0)	2048 (512)	1707 (853)	1463 (1097)	1280 (1280)

Memoryless Gauss white noise channel and BPSK modulation are adopted. Simulation result on MATLAB is shown in Fig. 2. After analysis, the algorithm proposed in this paper has obvious improvement on irregular LDPC codes.

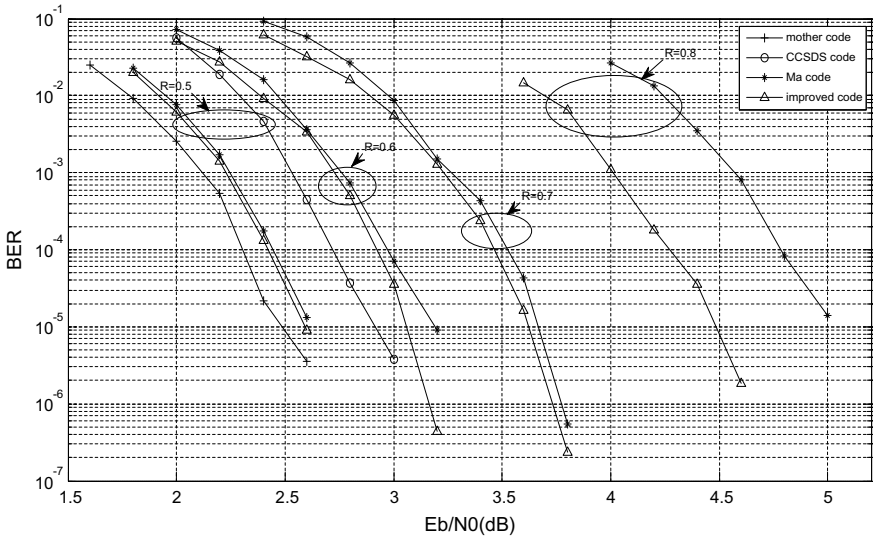


Fig. 2 Comparison between the improved algorithm and centralized puncturing algorithm; the mother code at rate 0.4 (leftmost) has a block length of 2560, and the CCSDS code at rate 0.5 has a block length of 2048, and the punctured LDPC codes have rates of 0.5, 0.6, 0.7, and 0.8 from left to right

At rate 0.5 and a BER of 10^{-5}, the punctured LDPC code under proposed improved algorithm has 0.35 dB better E_b/N_0 performance over CCSDS code, and 0.02 dB over Ma's code. At rate 0.8 and a BER of 10^{-5}, the proposed algorithm requires 0.5 dB less E_b/N_0 than that of centralized puncturing algorithm.

5 Conclusion

We propose the improved centralized puncturing algorithm to design rate-compatible punctured LDPC. The algorithm is based on the claim that a punctured LDPC code with a small level of recoverability has better performance.

The proposed algorithm is verified by comparing the performance of punctured LDPC codes based on the algorithm with unimproved algorithm. We apply the proposed algorithm to irregular LDPC codes at block length 2560. The performance improvements are 0.5 dB over punctured LDPC codes with Ma Fuli's algorithm at code rate 0.8 and a BER of 10^{-5}.

Acknowledgments. This work was supported by the Fundamental Research Funds for the Center Universities (Grant No. HIT.MKSTISP.2016 13).

References

1. Hagenauer, J.: Rate-compatible punctured convolutional codes(RCPC codes) and their applications. IEEE Trans. Commun. **36**, 389–400 (1988)
2. Babich, F., Montorsi, G., Vatts, F.: Some notes on rate-compatible punctured turbo codes (RCPTC). IEEE Trans. Commun. **52**, 681–684 (2004)
3. Li, J., Narayanan, K.: Rate-compatible low density parity check codes for capacity approaching ARQ scheme in packet data communication. In: Proceedings of International Conference on Communications, Internet, and Information Technology (CIIT) (2002)
4. MacKay, D.J.C., Neal, R.M.: Near Shannon limit performance of low-density parity check codes. Electron. Lett. **32**, 1645–1646 (1966)
5. Gallager, R.G.: Low density parity check codes. IEEE Trans. Inf. Theory **8**(3), 208–220 (1962)
6. Tanner, R.M.: A recursive approach to low complexity codes. IEEE Trans. Inf. Theory, IT-**27** (9), 533–547 (1981)
7. Ha, J., Kim, J., Klinc, D., et al.: Rate-compatible punctured low-density parity-check codes with short block lengths. IEEE Trans. Inf. Theory **52**(2), 728–738 (2006)
8. Park, H.Y., Kang, J.W., Kim, K.S., et al.: Efficient puncturing method for rate-compatible low-density parity-check codes. IEEE Trans. Wireless Commun. **6**(11), 3914–3919 (2007)
9. Fuli, M.: Puncturing schemes for rate-compatible LDPC codes. Beijing Jiaotong University (2012)

Polar Code Construction for BP Decoder

Xuan Yi, Aijun Liu, Qingshuang Zhang, Xiaohu Liang(✉), and Zhiyong Chen

Army Engineering University of PLA, Nanjing 210007, China
yixuan51321@163.com, liangxiaohu688@163.com

Abstract. Belief propagation (BP) is a high-throughput decoding algorithm for polar codes, but the performance under BP decoder is not satisfactory due to the mismatch between the virtual channels seen by the BP decoder and bit-channel in the conventional construction method. In this letter, we record the required number of iteration of unfrozen bits to reach a specific log-likelihood ratio (LLR), which can identify weak bit-channels. Moreover, we modify the conventional polar code construction by swapping these bit-channels with strong frozen bit-channels. Simulation results show that the proposed method achieves better performance than that of the conventional construction.

Keywords: Polar codes · Belief propagation decoding · Iteration

1 Introduction

Based on channel polarization, polar codes are introduced by Arikan in [1], whose capacity-achieving property and explicit structure have attracted the attention of many researchers in recent years. The first decoding algorithm, named successive-cancelation (SC), is proposed by Arikan [1]. However, SC performs quite poor when the length of polar codes is finite. To further improve the finite-length performance of polar codes, several decoders have been proposed. In [2], the performance of successive-cancelation list (SCL) decoding is comparable to maximum-likelihood (ML) decoding at high signal-to-noise ratio (SNR) by maintaining L candidate paths in the code tree concurrently. But it also brings higher computational complexity and latency. Belief propagation (BP) decoding over the polar code factor graph was also proposed, with parallel [3] and sequential [4] message scheduling. The parallel BP decoder provides a higher throughput and a reduced latency. However, the error-correcting performance of BP decoding is unsatisfying.

As to polar code construction, several calculation-based algorithms have been proposed, such as the density evolution (DE) [5] and Gaussian approximation (GA) algorithms [6,7]. These algorithms are all suitable for SC decoding. But for BP decoder, if we use these construction methods, it may not lead to the optimal performance under the BP decoder. Because the virtual channels between the input sequence to a linear encoder and the channel output sequence seen by a SC decoder may be different from the BP decoder [8].

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 232–240, 2019.
https://doi.org/10.1007/978-981-13-6264-4_28

In order to improve BP decoders, modified constructions of polar codes have been considered. For BP SCAN decoder, [8] uses the evolution of messages of unfrozen bits during iterative BP decoding of polar codes to identify weak bit-channels, and then modifies the conventional polar code construction by swapping these bit-channels with strong frozen bit-channels. While for flooding BP decoder, [9] presents a frozen bit selection algorithm for polar codes based on Monte Carlo simulation, where the information bits are selected out one by one incrementally. In [10], in-order bit selection method followed by the derivation of channel polarization is proposed. The two simulation-based constructions method bring a higher complexity.

In this work, a iteration-based construction method is proposed to further improve the performance of polar codes under BP decoder. By recording required iteration of unfrozen bits to reach a specific LLR during BP decoding, we find the unsuitable unfrozen bit-channels for the BP decoder which are then replaced by a partial frozen set one by one. Compared to the LLR-oriented construction, our method yields performance improvements under BP decoding.

2 Preliminaries

Following the notation of [1], we write W : X$\longrightarrow$Y to denote a binary-input discrete and memoryless channel with input $x \epsilon$ X, output $y \epsilon$ Y, and transition probabilities $W(y|x)$. We write $u_1^N = (u_1, u_2, ..., u_N)$ to denote the sequence of input bits and $x_1^N = (x_1, x_2, ..., x_N)$ to denote the sequence of coded bits, and N is the length of the code.

2.1 Polar Codes

Based on channel polarization [1], for a polar code with the code length $N = 2^n, n = 1, 2...$, given the code rate $R = K/N$, we can construct a polar code by selecting K good channels as nonfrozen channels which carry information bits, while the remaining $N - K$ channels are used to transmit some known values, usually zero. The set of frozen channel indices is denoted by A^c and the set of nonfrozen channel indices is denoted by A. The sequence of input bits $u_1^N = (u_1, u_2, ..., u_N)$ consists of K information bits u_A and $N - K$ frozen bits u_{A^c}. Accordingly, polar codewords $x_1^N = (x_1, x_2, ..., x_N)$ can be obtained as follows:

$$x_1^N = u_1^N \bullet G_N \tag{1}$$

where $G_N = F^{\otimes n}$ is generation matrix, $F^{\otimes n}$ is the $n{-}th$ Kronecker power of F, and

$$F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \tag{2}$$

is called the kernel matrix.

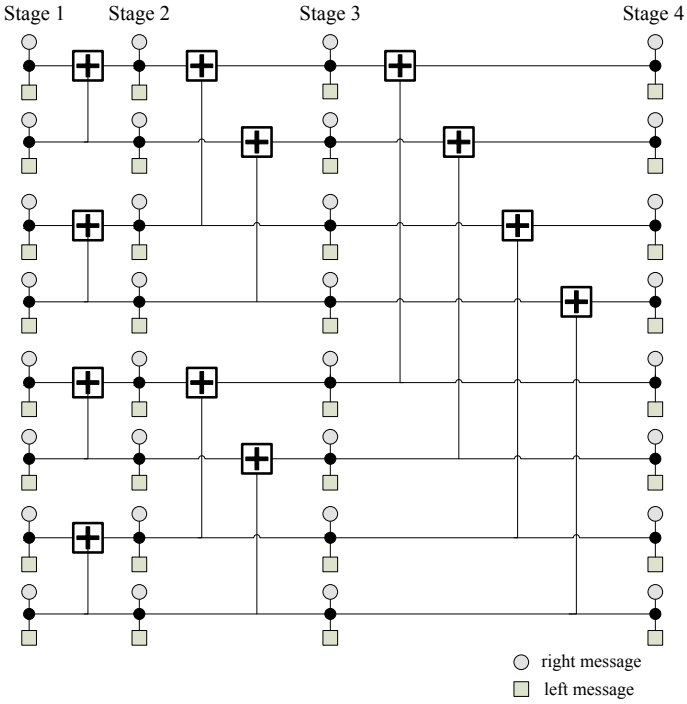


Fig. 1. Factor graph of the polar code with $N = 8$.

2.2 BP Decoding of Polar Codes

BP is a message-passing algorithm operating on a factor graph. As shown in Fig. 1, the BP decoder updates two types of messages of each node: left-to-right (left) messages and right-to-left (right) messages [11,12]. First, the right messages in stage 1 are initialized according to the information set A

$$R_{1,j} = \begin{cases} 0 \text{ if } j \in A \\ \infty \text{ if } j \in A^C \end{cases} \tag{3}$$

and the right messages in other stages are initialized to zero. For the left message, the initialization of nodes in stage $n+1$ are the channel output log-likelihood ratios (LLRs)

$$L_{n+1,j} = \ln \frac{P\left(y_j | x_j = 0\right)}{P\left(y_j | x_j = 1\right)} = \frac{2y_i}{\sigma^2} \tag{4}$$

and the left message rest nodes are initiated as zero.

Then, based on factor graph, the right messages update from stage n to stage 1 and left messages inversely update in each iteration. In the iteration, the

messages of each node are updated by

$$\begin{aligned} L_{i,j} &= g\left(L_{i+1,2j-1}, L_{i+1,2j} + R_{i,j+N/2}\right) \\ L_{i,j+N/2} &= g\left(R_{i,j}, L_{i+1,2j-1}\right) + L_{i+1,2j} \\ R_{i+1,2j-1} &= g\left(R_{i,j}, L_{i+1,2j} + R_{i,j+N/2}\right) \\ R_{i+1,2j} &= g\left(R_{i,j}, L_{i+1,2j-1}\right) + R_{i,j+N/2} \end{aligned} \tag{5}$$

where $g(x, y) = \log(cosh((x + y)/2)) - log(cosh((x - y)/2))$. In order to reduce the complexity of polar BP decoding, [12] proposes a scaling parameter α for approximation. the scaled min-sum (SMS) approximation can achieve decoding performance which is similar to that of the original BP algorithms.

After a specified number I_{max} of iterations, the estimate of u_1^N can be obtained by

$$\hat{u}_j = \begin{cases} 0 \text{ if } R_{1,j} + L_{1,j} \geq 0 \\ 1 \text{ if } R_{1,j} + L_{1,j} < 0 \end{cases} \tag{6}$$

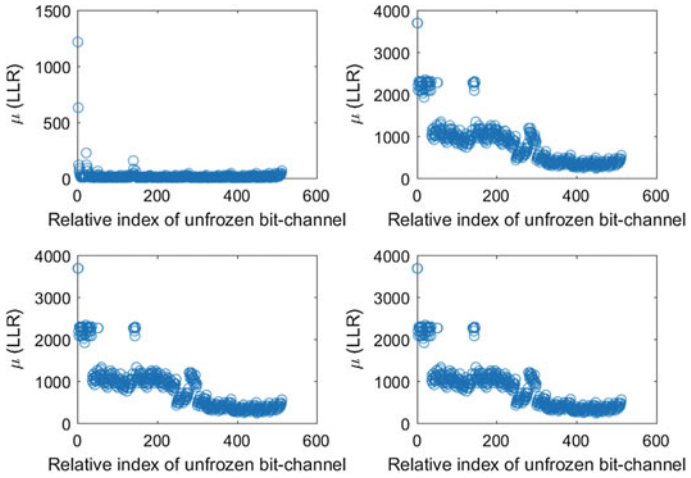


Fig. 2. Average values of LLRs $\mu(LLR)$ under BP decoding of unfrozen bits of polar code with $N = 1024$.

3 Iteration-Based Construction Algorithm

3.1 Analysis of Iteration During BP Decoding

For a BP decoder, the soft-output LLRs stabilizes gradually along with iterative process. Assume that an all-zero codeword is transmitted using binary phase-shift keying (BPSK) modulation. As shown in Fig. 2 with 5000 transmission of (1024, 512) code, after some iterations, about all average values of LLRs tend to be stable. The average values of LLRs of different bit-channel under stable state

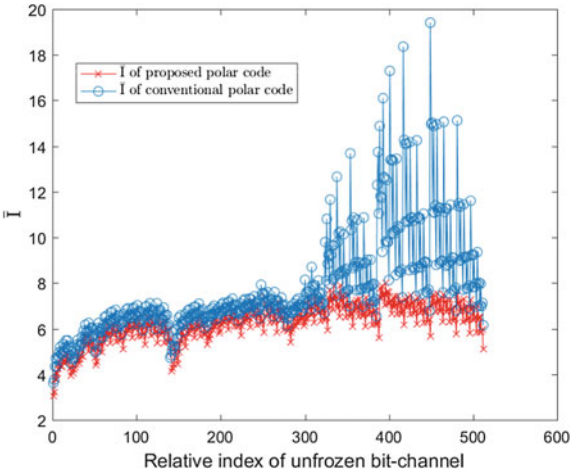


Fig. 3. $\bar{\mathrm{I}}$ under BP decoding of unfrozen bits of conventional polar code and proposed polar code with $LLR_{rth} = 6$, on AWGN channel, with $N = 1024$, $R = 0.5$ and $E_b/N_0 = 2.25\,\mathrm{dB}$.

is different. Some bit-channels have higher LLR while some near 0. According to formula (6), the decision threshold is 0. In order to successfully decode, for every information bit-channels, their LLR only need to be greater than a nonnegative number LLR_{rth}. LLR_{rth} is a parameter we set in our method. For a specific LLR_{rth}, for the bit-channel with higher soft decisions LLR, it usually takes only a small number of iterations to reach it. While for someone whose soft decisions LLR near 0, it need more iterations to reach it. Sometimes it fails to reach LLR_{rth} even though the BP decoder agrees over consecutive iterations. Intuitively, the error usually happens at these bit-channels, defined as weak bit-channel in our method.

Figure 3 shows the average required numbers of iteration to reach a specific LLR of BP decoding for a conventional polar code with length $N = 1024$ and rate $R = 0.5$, optimized at $E_b/N_0 = 2.25\,\mathrm{dB}$ [13]. Note that the plot shows average required iterations for the 512 unfrozen bits, displayed in corresponding bit-channel index order. For some bit-channels, the required number of iteration $\bar{I}$ is so high that the soft decisions LLR will be probable to lower than 0 after I_{max} of iterations and it leads to decoding fail.

3.2 Bit-Swapping Construction

In this section, we propose a modified polar code construction. In the simulations above, the error is more likely to occur at weak bit-channels. Therefore, we replace these weak nonfrozen bit-channels by an equal number of the frozen bit-channels from a partial frozen set A_p^c, whose size is determined empirically. This method is summarized in Algorithm 1, where $I_{\max}^p$ is the maximum of $\bar{\mathrm{I}}$ in Monte

Carlo simulation result with the previous nonfrozen set: A_{It}, and the set A_p^c is the subset of frozen set A^c with the order from small Bhattacharyya parameter to the large one.

In the Algorithm 1, we first run Monte Carlo BP decoding simulations with origin nonfrozen set A and record the required numbers of iteration $\bar{\mathrm{I}}$ to reach a LLR_{rth}, and then use the first element of set A_p^c to replace the bit-channnel with the maximal iteration $\bar{I}$. In this way, we can obtain a new nonfrozen set A_{It}. Similarly, we run Monte Carlo BP decoding simulations with new nonfrozen set A_{It} and conduct the swap. If the required numbers of iteration $\bar{\mathrm{I}}$ with new nonfrozen set A_{It} is satisfied, the swap is successful, otherwise, swap the next first element in the set A_p^c, corresponding to lines 4 to 12 in the Algorithm 1. After each swap, we can do the same conduction and obtain another new nonfrozen set A_{It} according to the rule mentioned above. Algorithm 1 ends when all elements of A_p^c have been swapped. The red solid line in Fig. 1 shows the distribution of $\bar{\mathbf{D}}$ using Iteration-based construction with $LLR_{rth} = 6$.

Algorithm 1: Iteration-based Construction

Require: nonfrozen set with conventional construction: A,
partial frozen set: A_p^c,
threshold of soft-output LLRs: LLR_{rth},
The number of Monte Carlo runs: $Nrun \leftarrow 5000$;
Ensure: nonfrozen set with Iteration-based
Construction: A_{It};
1: Initialization: $Len \leftarrow$ size of set A_p^c, $A_{It} \leftarrow A$, $I_{\max}^p \leftarrow I_{max}$;
2: **while** $Len > 0$ **do**
3: Run Monte Carlo BP decoding simulations with nonfrozen set A_{It} and record the required numbers of iteration to reach a LLR_{rth};
4: **if** $\max(\bar{\mathrm{I}}) \leq I_{\max}^p$ **then**
5: find the bit-channel B_m with the max number of iteration $\bar{I}$;
6: swap the bit-channel B_m with the first element of set A_p^c;
7: $B_{tmp} \leftarrow$ the first element of set A_p^c and remove the first element out of set A_p^c;
8: $I_{\max}^p \leftarrow \max(\bar{\mathrm{I}})$;
9: **else**
10: find the B_t in set A_p^c;
11: swap the bit-channel B_t with the first element of set A_p^c;
12: $B_t \leftarrow$ the first element of set A_p^c and remove the first element out of set A_p^c;
13: **end if**
14: $Len \leftarrow$ size of set A_p^c;
15: **end while**

4 Simulation Result

In this section, we will give some simulation results to verify that our method can improve the performance of polar codes. AWGN channel is assumed and the conventional construction used to construct polar codes is at $\sigma^2 = 0.49$.

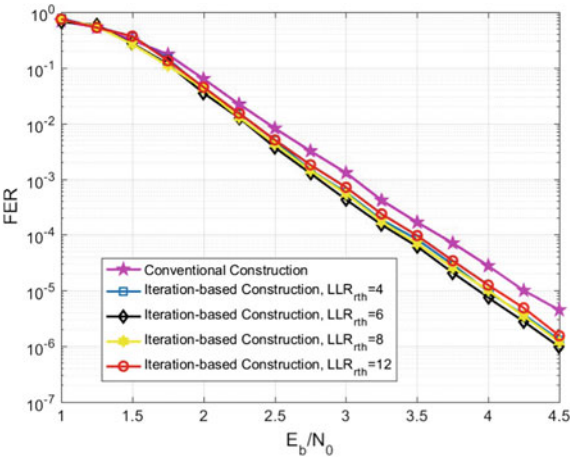


Fig. 4. Comparisons of FER performance between conventional construction and Iteration-based construction with different LLR_{rth}, for the transmission of (1024, 512) code under BP decoding, where $\sigma^2 = 0.49$.

For a proper comparison, we consider various code lengths (1024, 512) and code rates (1/3, 1/2, 2/3). At the receiver side, we employ SMS BP polar code decoder with scale parameters $\alpha = 0.9375$ and $I_{max} = 40$. In Fig. 4, for the target error probability $P_e = 10^{-3}$, an improvement of 0.25 dB with respect to the original polar code can be noticed using the Iteration-based construction with $LLR_{rth} = 6$. The finite-length performance of polar codes is significantly improved by using the Iteration-based construction for a suitable choice of the parameter LLR_{rth}. Specially, if the FER is decreasing, the improvement trend is increasing, e.g., when the FER is at 10^{-5}, the proposed decoder achieves an improvement of 0.4 dB.

Figure 5 gives the comparison of FER between the conventional construction and Iteration-based construction with the most empirically suitable LLR_{rth} for (1024, 341), (1024, 512), and (1024, 683) code. As shown, the improvement is more remarkable for higher code rate. For (1024, 512) and (1024,683) code, the proposed strategy creates 0.25 dB and 0.5 dB improvement with the target error probability $P_e = 10^{-3}$. While for (1024, 341) code, the FER slightly decreases compared to the conventional one, this is because the FER performance is mainly influenced by the incompletely polarized bit-channels. That is to say, for a specific higher rate, all nonfrozen bits are more suitable for the BP decoder than other frozen bit-channels and there is no need to swap them with the frozen set.

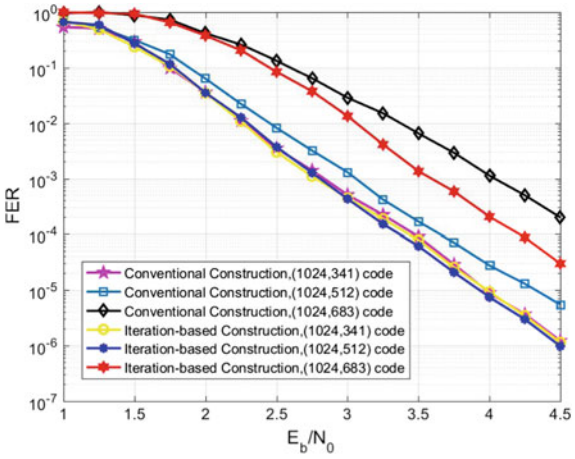


Fig. 5. Comparisons of FER performance between Iteration-based construction and conventional construction for different code rate.

5 Conclusion

In this work, we proposed a simulation-based bit-swapping construction method of polar codes for BP decoding. With Monte Carlo simulation, our algorithm finds that some weak nonfrozen bits need to be replaced, and then we swap them with the same number of the most reliable frozen bit-channels. Simulation results indicate that the Iteration-based construction exhibit significantly better performance than the conventional codes.

References

1. Arikan, E.: Channel polarization: a method for constructing capacityachieving codes for symmetric binary-input memoryless channels. IEEE Trans. Inf. Theory **55**(7), 3501–3573 (2009)
2. Tal, I., Vardy, A.: List decoding of polar codes. In: Proceedings of 2011 IEEE International Symposium on Information Theory, Austin, TX, USA, pp. 1–5, June 2011
3. Arikan, E.: A performance comparison of polar codes and Reed Muller codes. IEEE Commun. Lett. **12**(6), 447–449 (2008)
4. Fayyaz, U.U., Barry, J.R.: Polar codes for partial response channels. In: Proceedings of IEEE International Conference on Communication (ICC), pp. 4337–4341, June 2013
5. Mori, R., Tanaka, T.: Performance of polar codes with construtions using density evolution. IEEE Commun. Lett. **13**(7), 192–203 (2009)
6. Trifonov, P.: Efficient design and decoding of polar codes. IEEE Trans. Commun. **60**(11), 3221–3227 (2012)
7. Wu, D., Li, Y., Sun, Y.: Construction and block error rate analysis of polar codes over AWGN channel based on Gaussian approximation. IEEE Commun. Lett. **18**(7), 1099–1102 (2014)

8. Qin, M.H., Guo, J., Bhatia, A., Fabregas, A., Siegel, P.H.: Polar code constructions based on LLR evolution. IEEE Commun. Lett. **21**(6), 1221–1224 (2017)
9. Liu, J.B., Sha, J.: Frozen bits selection for polar codes based on simulation and BP decoding. In: IEICE Electronics Express, vol. 14, no. 6 (2017)
10. Sun, S.H., Zhang, Z.: Design practical polar codes using simulation-based bit selection. IEEE J. Emerg. Sel. Topics Circ. Syst. **7**(4), 594–603 (2017)
11. Eslami, A., Pishro-Nik, H.: On finite-length performance of polar codes: stopping sets, error floor, concatenated design. IEEE Trans. Commun. **61**(3), 919–929 (2013)
12. Yuan, B., Parhi, K.K.: Architecture optimizations for BP polar decoders. In: Proceedings of 38th IEEE ICASSP, pp. 2654–2658, May 2013
13. Tal, I., Vardy, A.: How to construct polar codes. IEEE Trans. Inf. Theory **59**(10), 6562–6582 (2013)

Robust Broadband Adaptive Beamforming with Constant Beamwidth for Moving Interference Suppression

Shuai Li, Xiaopeng Yang(✉), Junqi Xue, and Xuchen Wu

School of Information and Electronics, Beijing Institute of Technology, and Key Laboratory of Electronic and Information Technology in Satellite Navigation, Ministry of Education, Beijing 100081, China
xiaopengyang@bit.edu.cn

Abstract. The performance of the conventional broadband adaptive beamforming methods degrades severely in scenarios with moving interference. Moreover, due to the inconsistency of mainlobes width at different frequencies, the conventional methods also deteriorate when the target direction differs with the look direction. Therefore, a robust broadband adaptive beamforming method with constant beamwidth for moving interference suppression is proposed. In the proposed method, a taper matrix is constructed by adding two virtual interferences arranged at the two sides of each original interference. Then the array covariance matrix is reconstructed based on the taper matrix and the sample covariance matrix. Meanwhile, the spatial response variation (SRV) constraint is applied to guarantee a constant beamwidth over the entire frequency band. Finally, the adaptive weight vector for null broadening and constant beamwidth is calculated by the Lagrange multiplier method. The effectiveness of the proposed method has been verified by numerical simulations.

Keywords: Robust broadband adaptive beamforming · Moving interference · Null broadening · Constant beamwidth

1 Introduction

During the past three decades, the narrowband adaptive beamforming has been extensively investigated [1]. To obtain better performance, the broadband beamforming has attracted extensive attention. However, the conventional broadband adaptive beamforming methods are only suitable for the static interferences. The performance of the conventional methods degrades severely in the rapidly moving interferences environment since the interferences motion may bring the interferences out of the sharp nulls of the adaptive beampattern. Moreover, the mainlobe width of each frequency sub-band is inconsistent in the conventional broadband beamforming methods, resulting in the severe distortion of the target signal when the target direction differs from the looking direction. For the inconsistency of the mainlobe width, many methods have been investigated to achieve constant beamwidth over the whole frequency band, such as the Bessel function method [2] and the method with frequency invariance

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 241–248, 2019.
https://doi.org/10.1007/978-981-13-6264-4_29

(FI) constraints [3]. However, the method in [2] can only satisfy constant beamwidth, but it cannot suppress interferences effectively. The method in [3] can achieve constant beamwidth and form sharp nulls for the interferences, whereas the performance degrades severely in the case of moving interferences. In order to suppress the moving interferences robustly, the covariance matrix tapers method [4] and the sidelobe derivative constraints method are proposed. Although they can broaden the nulls for the interferences in the narrowband beamforming, they cannot be applied to broadband beamforming directly. Recently, several null broadening methods for broadband beamforming have been proposed [5, 6]. However, the method in [5] cannot form sufficiently wide nulls for interferences. The proposed method in [6] can suppress the moving interferences effectively, whereas it requires the interference direction as prior.

In this paper, a robust broadband adaptive beamforming method with constant beamwidth for moving interference suppression is proposed. Specifically, a taper matrix is calculated by adding two virtual interferences around each original interference, and then the array covariance matrix is reconstructed as the Hadamard product of the taper matrix and the sample covariance matrix. Meanwhile, the constant beamwidth over the entire desired frequency band is realized based on the SRV constraint. Eventually, the adaptive weight vector for null broadening and constant beamwidth is calculated by the Lagrange multiplier method.

The rest of this paper is organized as follows: the broadband signal model is introduced in Sect. 2. The proposed robust broadband adaptive beamforming method is described in Sect. 3. Simulation results and analyses are presented in Sect. 4. Finally, the conclusions are summarized in Sect. 5.

2 Broadband Signal Model

Consider a uniform linear array with M elements, N taps with each sensor, and P broadband interferences from directions $\theta_1, \cdots, \theta_P$ with a bandwidth B. The received data at the mth sensor is written as

$$x_m(t) = \sum_{i=1}^{P} s_i(t - \tau_m(\theta_i)) + n_m(t) \tag{1}$$

where $s_i(\cdot)$ is the corresponding complex envelop of the ith interference, $n_m(t)$ is the additive white Gaussian noise at the mth sensor, and $\tau_m(\theta_i)$ is the propagation delay associated with the ith interference and the mth sensor.

The covariance matrix of the received signals is denoted as $\boldsymbol{R}_{xx} = E\{\boldsymbol{x}(t)\boldsymbol{x}^H(t)\}$, where $\boldsymbol{x}(t) = [x_1(t), \cdots, x_M(t), \cdots, x_M(t - (N-1)T_s)]^T$, $(\cdot)^T$ is the transpose, and $(\cdot)^H$ denotes the complex conjugate transpose.

Using the fast discrete Fourier transform, the broadband signals are decomposed into L nonoverlapping sub-bands, and the received data $\boldsymbol{x}(t)$ in the frequency domain is written as

$$\begin{aligned} \boldsymbol{X}(f_l) &= \boldsymbol{A}(f_l)\boldsymbol{S}(f_l) + \boldsymbol{N}(f_l) \quad l = 1, 2, \ldots, L \\ \boldsymbol{A}(f_l) &= [\boldsymbol{a}(f_l, \theta_1), \boldsymbol{a}(f_l, \theta_2), \ldots, \boldsymbol{a}(f_l, \theta_P)] \end{aligned} \tag{2}$$

where $\boldsymbol{S}(f_l) = [S_1(f_l), \cdots, S_P(f_l)]^T$ is the Fourier transform vector of the complex envelopes of the interferences and $\boldsymbol{N}(f_l)$ is the Fourier transform of the array noise. $\boldsymbol{a}(f_l, \theta_i) = \boldsymbol{a}_{Ts}(f_l) \otimes \boldsymbol{a}_{\tau_m}(f_l, \theta_i)$ is the $MN \times 1$ dimensional steering vector of the ith interference, $\otimes$ represents the Kronecker product and

$$\begin{aligned} \boldsymbol{a}_{Ts}(f_l) &= [1, e^{-j2\pi f_l T_s}, \cdots, e^{-j2\pi f_l (N-1) T_s}]^T \\ \boldsymbol{a}_{\tau_m}(f_l, \theta_i) &= [1, e^{-j2\pi f_l d \sin\theta_i / c}, \cdots, e^{-j2\pi f_l (M-1) d \sin\theta_i / c}]^T \end{aligned} \tag{3}$$

where T_s is the sampling period, d is the interval of sensors, and c is the wave propagation speed.

The array spatial response, as a function of the frequency f_l and the signal direction θ is written as

$$\boldsymbol{H}(f_l, \theta) = \sum_{m=1}^{M} \sum_{n=1}^{N} w_{m,n}^{*} e^{-j2\pi f_l (m-1) d \sin\theta / c} e^{-j2\pi f_l (n-1) T_s} \tag{4}$$

where $w_{m,n}$ is the weight coefficient of the mth sensor and nth tap.

The response in (4) can also be expressed in the following vector form

$$\boldsymbol{H}(f_l, \theta) = \boldsymbol{w}^H \boldsymbol{a}(f_l, \theta) \qquad 1 \le l \le L \tag{5}$$

where $\boldsymbol{w} = [w_{1,1}, \cdots, w_{M,1}, \cdots, w_{1,N}, \cdots, w_{M,N}]^T$ is the weight vector.

3 Robust Broadband Adaptive Beamforming Method

In order to achieve constant beamwidth and suppress the moving interferences, a robust broadband adaptive beamforming method with constant beamwidth for moving interference suppression is proposed. In the proposed method, a taper matrix is calculated by adding two virtual interferences around each original interference, and then the array covariance matrix is reconstructed. Moreover, the SRV constraint is used to achieve constant beamwidth. Finally, the weight vector is calculated using the Lagrange multiplier method. The detailed procedures of proposed method are as follows.

3.1 Matrix Taper Method for Null Broadening

Taper Matrix Calculation

The original array covariance matrix $\boldsymbol{R}_f$ in the frequency domain is denoted as

$$\boldsymbol{R}_f = \sum_{l=1}^{L} \boldsymbol{A}(f_l) \boldsymbol{R}_s(f_l) \boldsymbol{A}^H(f_l) + \boldsymbol{I} \sum_{l=1}^{L} \sigma_n^2(f_l) \tag{6}$$

where $\boldsymbol{R}_s(f_l) = E\{\boldsymbol{S}(f_l)\boldsymbol{S}^H(f_l)\}$ is the covariance matrix of the complex envelopes of interferences, $\sigma_n^2(f_l)$ is the noise power at the frequency f_l, and $\boldsymbol{I}$ is the $MN \times MN$ dimensional identity matrix.

In order to form wide null at interference direction, assume that there are two virtual interferences with equal strength arranged at the two sides of each original interference. Moreover, the maximum offset of the ith interference direction $\theta_i (i = 1, \cdots, P)$ caused by the interference motion is denoted as $\Delta\theta$. Therefore, for the ith interference, the corresponding steering vector of each virtual interference is expressed as

$$\begin{aligned}\boldsymbol{a}(\theta_i + p\Delta\theta, f_l) =& \left[1, \cdots, e^{-j2\pi f_l (N-1)T_s}\right]^T \otimes \left[1, \cdots, e^{-j(M-1)\varphi_1}\right]^T \\ =& \left[1, e^{-j\varphi_1}, \cdots, e^{-j(M-1)\varphi_1}, e^{-j2\pi f_l T_s}, e^{-j\varphi_1} e^{-j2\pi f_l T_s}, \cdots, e^{-j(M-1)\varphi_1} e^{-j2\pi f_l (N-1)T_s}\right]^T\end{aligned} \tag{7}$$

where $p = \pm 1$, and φ_1 is denoted as $\varphi_1 = 2\pi d f_l \sin(\theta_i + p\Delta\theta)/c$.

When $\Delta\theta$ is a small value, we can obtain

$$sin(\theta_i + p\Delta\theta) = (\sin\theta_i \cos p\Delta\theta + \sin p\Delta\theta \cos\theta_i) \approx \sin\theta_i + p\Delta\theta\cos\theta_i \tag{8}$$

To obtain as large an angle offset as possible, we set $\cos\theta_i = 1$. Then, φ_1 is approximated as

$$\varphi_1 \approx 2\pi d f_l (\sin\theta_i + p\Delta\theta)/c \tag{9}$$

Substituting (9) into (7), we can obtain

$$\begin{aligned}\boldsymbol{a}(\theta_i + p\Delta\theta, f_l) = [1, e^{-j(\varphi + p\Delta\varphi)}, \cdots, e^{-j(M-1)(\varphi + p\Delta\varphi)}, \cdots, \\ e^{-j(M-1)(\varphi + p\Delta\varphi)} e^{-j2\pi f_l (N-1)T_s}]^T = \boldsymbol{Q}_p \boldsymbol{a}(\theta_i, f_l)\end{aligned} \tag{10}$$

where $\varphi = 2\pi d f_l \sin\theta_i / c$, $\Delta\varphi = 2\pi d f_l \Delta\theta / c$ and

$$\boldsymbol{Q}_p = \mathrm{diag}\{1, e^{-jp\Delta\varphi}, \cdots, e^{-j(M-1)p\Delta\varphi}, \cdots, 1, \cdots, e^{-j(M-1)p\Delta\varphi}\} \tag{11}$$

From (11), it is found $\boldsymbol{Q}_p \boldsymbol{Q}_p^H = \boldsymbol{I}$ and $\boldsymbol{Q}_{-1} = \boldsymbol{Q}_1^H$.

The array covariance matrix with additional virtual interferences is calculated as

$$\tilde{\boldsymbol{R}}_f = \boldsymbol{R}_f^r + \boldsymbol{R}_f + \boldsymbol{R}_f^l \approx \boldsymbol{Q}_1 \boldsymbol{R}_f \boldsymbol{Q}_1^H + \boldsymbol{R}_f + \boldsymbol{Q}_{-1} \boldsymbol{R}_f \boldsymbol{Q}_{-1}^H \tag{12}$$

where $\boldsymbol{R}_f^r$ and $\boldsymbol{R}_f^l$ represent the covariance matrices of the virtual interferences from $\theta_i + \Delta\theta$ and $\theta_i - \Delta\theta$, respectively. And they are expressed as

$$\begin{aligned}\boldsymbol{R}_f^r &\approx \boldsymbol{Q}_1 \sum_{l=1}^{L} \boldsymbol{A}(f_l) \boldsymbol{R}_s(f_l) \boldsymbol{A}^H(f_l) \boldsymbol{Q}_1^H + \boldsymbol{I} \sum_{l=1}^{L} \sigma_n^2(f_l) \\ \boldsymbol{R}_f^l &\approx \boldsymbol{Q}_{-1} \sum_{l=1}^{L} \boldsymbol{A}(f_l) \boldsymbol{R}_s(f_l) \boldsymbol{A}^H(f_l) \boldsymbol{Q}_{-1}^H + \boldsymbol{I} \sum_{l=1}^{L} \sigma_n^2(f_l)\end{aligned} \tag{13}$$

Further, the elements of $\boldsymbol{R}_f^r$ and $\boldsymbol{R}_f^l$ are denoted as

$$\begin{aligned}(\boldsymbol{R}_f^r)_{kl} &\approx r_{kl}e^{-j\frac{\Delta\varphi}{2}[\mathrm{rem}(k-1,M)-\mathrm{rem}(l-1,M)]} \\ (\boldsymbol{R}_f^l)_{kl} &\approx r_{kl}e^{j\frac{\Delta\varphi}{2}[\mathrm{rem}(k-1,M)-\mathrm{rem}(l-1,M)]} \quad 1 \le k,l \le MN\end{aligned} \tag{14}$$

where r_{kl} is the element of the original array covariance matrix $\boldsymbol{R}_f$, and rem[a, b] is the remainder after a divided b. Substituting (14) into (12), we obtain

$$(\tilde{\boldsymbol{R}}_f)_{kl} \approx r_{kl}\left(1+2\cos\left(\frac{\Delta\varphi}{2}[\mathrm{rem}(k-1,M)-\mathrm{rem}(l-1,M)]\right)\right) \tag{15}$$

Therefore, the taper matrix $\boldsymbol{T}$ which can broaden the interference null is defined as

$$\boldsymbol{T} = \boldsymbol{1}_{N\times N} \otimes \boldsymbol{T}_1 \tag{16}$$

where $\boldsymbol{1}_{N\times N}$ is a $N \times N$ dimensional all-ones matrix, and $\boldsymbol{T}_1$ is

$$(\boldsymbol{T}_1)_{gh} = 1+2\cos\left(\frac{\Delta\varphi}{2}[\mathrm{rem}(g-1,M)-\mathrm{rem}(h-1,M)]\right) \quad 1 \le g,h \le M \tag{17}$$

Array Covariance Matrix Reconstruction

The array covariance matrix can be reconstructed as the Hadamard product of the taper matrix and the sample covariance matrix, which is denoted as

$$\tilde{\boldsymbol{R}}_{xx} = \boldsymbol{R}_{xx} \odot \boldsymbol{T} \tag{18}$$

where $\odot$ denotes the Hadamard product.

3.2 SRV Constraint for Constant Beamwidth

Generally, the SRV is used to measure the fluctuation of the array spatial response over the entire desired frequency band [3]. In a general form, the SRV in the direction θ is expressed as

$$\mathrm{SRV}(\theta) = \frac{1}{B}\int\limits_{f\in F_C} \left|\boldsymbol{w}^H\boldsymbol{a}(f,\theta) - \boldsymbol{w}^H\boldsymbol{a}(f_r,\theta)\right|^2 \mathrm{d}f \tag{19}$$

where B is the bandwidth of the desired frequency range F_C, and f_r represents the reference frequency. Further, the average SRV over the specified angle set Θ_C is calculated as

$$\begin{aligned}\overline{\mathrm{SRV}} &= \frac{1}{B}\frac{1}{\Theta_C}\int\limits_{f\in F_C}\int\limits_{\theta\in\Theta_C}\left|\boldsymbol{w}^H(\boldsymbol{a}(f,\theta)-\boldsymbol{a}(f_r,\theta))\right|^2\mathrm{d}\theta\mathrm{d}f = \boldsymbol{w}^H\boldsymbol{V}\boldsymbol{w}\\ \boldsymbol{V} &= \frac{1}{B}\frac{1}{\Theta_C}\int\limits_{f\in F_C}\int\limits_{\theta\in\Theta_C}(\boldsymbol{a}(f,\theta)-\boldsymbol{a}(f_r,\theta))(\boldsymbol{a}(f,\theta)-\boldsymbol{a}(f_r,\theta))^H\mathrm{d}\theta\mathrm{d}f\end{aligned} \tag{20}$$

where $\overline{\mathrm{SRV}}$ represents the average SRV.

When the $\overline{\mathrm{SRV}}$ is constrained to a small positive term γ, the constant beamwidth over the bandwidth B and the angle set Θ_C can be obtained.

3.3 Adaptive Weight Vector Calculation

The adaptive weight vector is calculated by minimizing the sum of the output power of the beamformer and the average SRV, with the array response being constrained to the desired value D

$$\begin{aligned}\min_{\boldsymbol{w}} \quad & \boldsymbol{w}^H(\beta\tilde{\boldsymbol{R}}_{xx}+(1-\beta)\boldsymbol{V})\boldsymbol{w}\\ \text{subject to} \quad & \boldsymbol{w}^H\boldsymbol{a}(f_r,\theta_0)=\mathrm{D}\end{aligned} \tag{21}$$

where β is a trade-off parameter between the constant beamwidth property and the output power of the beamformer. θ_0 is the looking direction. The adaptive weight vector is calculated by using the Lagrange multiplier method, denoted as

$$\boldsymbol{w} = \frac{D(\beta\tilde{\boldsymbol{R}}_{xx}+(1-\beta)\boldsymbol{V})^{-1}\boldsymbol{a}(f_r,\theta_0)}{\boldsymbol{a}^H(f_r,\theta_0)(\beta\tilde{\boldsymbol{R}}_{xx}+(1-\beta)\boldsymbol{V})^{-1}\boldsymbol{a}(f_r,\theta_0)} \tag{22}$$

From (21) and (22), it is found that the proposed method can not only realize a constant beamwidth over the entire frequency band, but also suppress interferences robustly. Thus, the proposed method is effective for an actual broadband system.

4 Simulation Results

To evaluate the performance of the proposed method, numerical simulations have been carried out. In the simulations, a uniform linear array with 20 omnidirectional sensors and 21 taps is considered. The array spacing is half-wavelength corresponding to the highest frequency. The mainlobe direction is 0°, and the Gaussian white additive noise is considered. One interference is from 40° with the interference-to-noise ratio (INR) 40 dB. Assume that the received signals are all centered at the frequency of 1.3 GHz with a bandwidth of 200 MHz. Moreover, the desired response is D = 1, and the trade-off parameter is $\beta = 0.01$.

The array beampatterns of the Frost method, the FI constraints method in [3] and the proposed method are investigated, and the corresponding results are shown in Fig. 1. It is found that the mainlobe beamwidth obtained by the Frost method differs

along with the frequency bins, whereas the FI constraints method and the proposed method can achieve constant mainlobe beamwidth over the design frequency band.

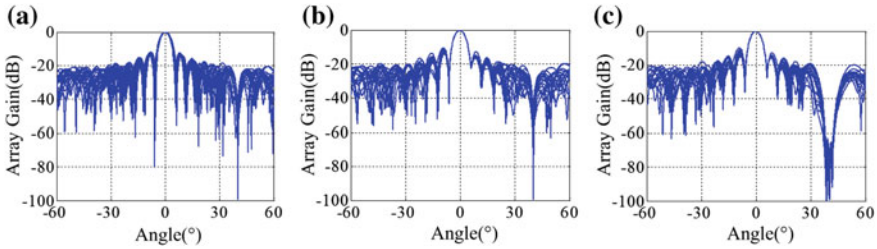


Fig. 1 Array beampatterns obtained by the Frost method shown in (a), the FI constraints method shown in (b), the proposed method shown in (c).

Moreover, the FI constraints method and the Frost method can only form a sharp null at the interference direction. However, the proposed method can form trough region centered at the interference direction, which is much wider than the results of other two methods.

Consider that there is one moving pulse interference and the interference direction of each pulse is different. The trajectory of the interference motion versus the pulse index k is denoted as $\theta(k) = 40^\circ + 2^\circ \sin[(k-1)/5] \quad 1 \le k \le N_p$. N_p = 60 is the total number of pulses. The output signal to interference and noise ratios (SINRs) versus the pulses for input signal-to-noise ratio (SNR) in each sensor fixed at 0 dB are investigated by the proposed method, the Optimum, the Frost method and the FI constraints

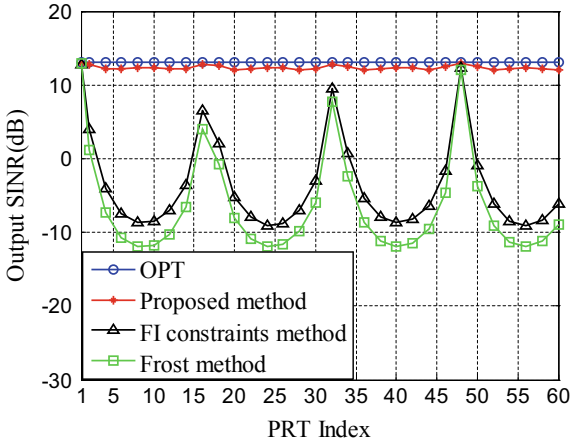


Fig. 2 Output SINRs versus the pulses obtained by the Optimum, the Frost method, the FI constraints method and the proposed method, respectively.

method and the results are shown in Fig. 2. Note that the adaptive weight vectors of the three methods are calculated by data in first pulse, then the weight vectors are used to calculate SINR of each pulse with 200 independent Monte Carlo simulations.

It is found that the output SINRs of the Frost method and the FI constraints method fluctuate in a sinusoidal pattern with pulse index. This is because the two methods can only form sharp nulls at direction of the interference in the first pulse. Once the direction of the interference changes, the interference will move out of the sharp nulls and the output SINRs of the two methods degrade severely. However, the output SINRs obtained by proposed method are almost constant, and the SINR loss is approximately only 0.5 dB relative to the optimum value.

5 Conclusions

In this paper, a robust broadband adaptive beamforming method with constant beamwidth for moving interferences suppression is proposed. In the proposed method, the taper matrix is calculated by adding two virtual interferences around each original interference. Then, the covariance matrix is reconstructed based on the taper matrix and the sample covariance matrix. Meanwhile, the constant beamwidth over the desired frequency band is achieved based on the SRV constraint. Eventually, the adaptive weight vector for null broadening and the constant beamwidth is calculated by using the Lagrange multiplier method. Simulation results show that the proposed method can realize the constant beamwidth and form wide nulls for interferences. Therefore, the proposed method is robust for a practical broadband system.

Acknowledgement. This work was supported by National Natural Science Foundation of China (Grant Nos. 61860206012, 61671065 and 31727901), and 111 Project of China (Grant No. B14010)

References

1. Chen, X., Shu, T., Yu, K.B., Yu, W.: Enhanced ADBF architecture for monopulse angle estimation in multiple jammings. IEEE Antennas Wirel. Propag. Lett. **16**, 2684–2687 (2017)
2. Yang, Y., Sun, C.: Arbitrary array structure constant beamwidth beamforming methods. Acta Acust. **26**(1), 55–58 (2001)
3. Zhao, Y., Liu, W., Langley, R.J.: Adaptive wideband beamforming with frequency invariance constraints. IEEE Trans. Antennas Propag. **59**(4), 1175–1184 (2011)
4. Guerci, J.R.: Theory and application of covariance matrix tapers for robust adaptive beamforming. IEEE Trans. Signal Process. **47**(4), 977–985 (1999)
5. Ren, C., Wu, S., Wang, J., Li, J.: A robust adaptive beamforming algorithm based on space-time processing. J. Electron. Inf. Technol. **31**(6), 1381–1385 (2009)
6. Li, S., Yang, X., Ning, L., Long, T., Sarkar, K.T.: Broadband constant beamwidth beamforming for suppressing mainlobe and sidelobe interferences. In: IEEE Radar Conference, pp. 1041–1045 (2017)

Outage Probability Analysis of Decode-and-Forward Based Hybrid Satellite–Terrestrial Relay Networks

Guoqiang Cheng, Qingquan Huang(✉), Xiangshuai Tao, Lin Yang, and Min Lin

Army Engineering University of PLA, Nanjing, China
huangqingquan398@163.com

Abstract. This paper investigates the performance of a hybrid satellite–terrestrial cooperative relay network with the relay node equipped N antennas, where the satellite-destination and satellite-relay links follow the shadowed-Rician fading, and the relay-destination link undergoes the correlated Rayleigh fading. First, we derive the expression for probability density functions of signal-to-noise ratio in satellite-destination and satellite-relay links by applying maximum ratio combining protocol at the satellite, then the expression in relay-destination link is derived by employing the maximum ratio transmission at the relay. Next, we obtain the expression of outage probability of this system. Finally, the numerical results are given to validate the theoretical analysis, and we find that the performance of this system gets better with the increasing number of relay's antennas.

Keywords: Multi-antennas · Satellite communication · Correlated Rayleigh fading channels

1 Introduction

Satellite communication systems are widely used for providing service over a broad coverage area. However, the line of sight (LOS) between satellite and user is always blocked by the masking effect due to shadowing and obstacles. As we know, the integration of relay technique into satellite communications has been regarded as an effective way to remit the severe fading of the channel. Under this situation, the hybrid satellite–terrestrial cooperative networks (HSTCNs) have gained more attention due to their ability to achieve reliable transmission and broad coverage [1].

Many workers have focused on the performance evaluation of HSTCNs thus far. In [2], the outage probability of HSTCN with best relay has been derived. The author in [3] analyzed the performance of amplify-and-forward (AF) [4] based HSTCN over generalized fading channels, and analytical diversity order of the hybrid system is also obtained. Moreover, in [5], the author investigated the performance of HSTCN with distributed Alamouti Code by assuming that AF protocol is adopted at the terrestrial relay.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 249–256, 2019.
https://doi.org/10.1007/978-981-13-6264-4_30

Although the foregoing workers have analyzed the performance of cooperative network, they have not researched the performance of a HSTCN with a multi-antennas relay and the terrestrial channel undergoing correlated Rayleigh fading. To fill this gap, we consider a cooperative network with multi-antennas relay which implements decode-and-forward (DF) protocol, where the Relay-Destination (R-D) link follows correlated Rayleigh fading. By applying maximum ratio combining (MRC) and maximum ratio transmission (MRT) protocols at the relay, and selection combining protocol at D, we analyze the outage probability (OP) of this network. Finally, numerical results are provided to validate the correctness of the expression of OP. To the best of our knowledge, this is the first time that such expression has been derived.

2 System Model

As illustrated in Fig. 1, we consider a HSTCN where the satellite (S) communicates with a terrestrial user (D) via a terrestrial relay (R). We assume that the satellite and the terrestrial destination are both equipped with single antenna and the terrestrial relay node is equipped with N antennas. The S-D and the S-R links undergo the shadowed-Rician fading while the R-D link undergoes the correlated Rayleigh fading.

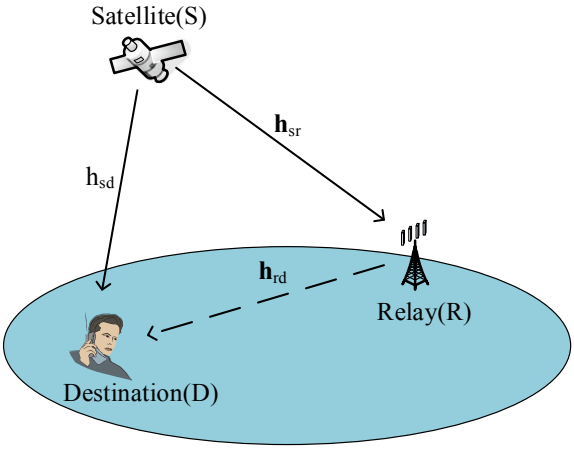


Fig. 1. System model

The overall communication is divided into two time slots. During the first time slot, the satellite transmits signal $x_s(t)$ with $\mathrm{E}[|x_s(t)|^2] = 1$ to the relay and destination, then the received signals at R and D can be respectively written as

$$y_{sd}(t) = \sqrt{P_s} h_{sd}(t) x_s(t) + n_{sd} \tag{1}$$

$$y_{sr}(t) = \sqrt{P_s} \mathbf{h}_{sr}^H \mathbf{w}_{sr} x_s(t) + n_{sr} \tag{2}$$

where $\mathbf{w}_{sr} = [w_{sr_1}, w_{sr_2} \ldots w_{sr_N}]^T$ denotes the receive beamforming weight vector at R, P_s denotes the transmit power at S, h_{sd} and $\mathbf{h}_{sr}$ the channel coefficient of S-D link and S-R link, respectively, with $\mathbf{h}_{sr} = [h_1, h_2, \cdots, h_N]^T$. Meanwhile, the additive white Gaussian noise (AWGN) at D and R are represented as $n_{sd} \sim N_C(0, \sigma_0^2)$ and $n_{sr} \sim N_C(0, \sigma_1^2)$. Thus, the output SNRs at D and R can be, respectively, given by

$$\gamma_{sr} = \frac{P_s}{\sigma_0^2} |h_{sd}|^2 \tag{3}$$

$$\gamma_{sr} = \frac{P_s}{\sigma_1^2} \mathbf{w}_{sr}^H \mathbf{h}_{sr} \mathbf{h}_{sr}^H \mathbf{w}_{sr} \tag{4}$$

During the second slot, R first decodes the received signal $y_{sr}(t)$, then a re-encoded signal $x_r(t)$ with an average power $\mathrm{E}[|x_r(t)|^2] = 1$ is transmitted from R to D by using DF protocol, the received signal of the relay link at D is given by

$$y_{rd}(t) = \sqrt{P_r} \mathbf{h}_{rd}^H \mathbf{w}_{rd} x_r(t) + n_2 \tag{5}$$

where $\mathbf{w}_{rd} = [w_{rd_1}, w_{rd_2}, \ldots, w_{rd_N}]^T$ and $n_2 \sim N_C(0, \sigma_2^2)$. As such, we can get the SNR of the R-D link

$$\gamma_{rd} = \frac{P_r}{\sigma_2^2} \mathbf{w}_{rd}^H \mathbf{h}_{rd} \mathbf{h}_{rd}^H \mathbf{w}_{rd} \tag{6}$$

Since DF protocol is applied at R, the SNR of S-R-D link can be expressed as [2]

$$\gamma_{df} = \min\{\gamma_{sr}, \gamma_{rd}\} \tag{7}$$

Finally, by applying selection combining at D, the instantaneous output signal-to-noise ratio (SNR) can be expressed as

$$\gamma = \max\left\{\gamma_{df}, \gamma_{sd}\right\} = \max\left\{\min\{\gamma_{sr}, \gamma_{rd}\}, \gamma_{sd}\right\} \tag{8}$$

3 Statistical Analysis

3.1 Satellite Channel

In order to get the maximum SNR of S-R link, such that

$$\max_{\mathbf{w}_{sr}} \gamma_{sr} = \frac{P_s}{\sigma_1^2} \mathbf{w}_{sr}^H \mathbf{h}_{sr} \mathbf{h}_{sr}^H \mathbf{w}_{sr}, \quad s.t. ||\mathbf{w}_{sr}||^2 = 1 \tag{9}$$

as $\text{rank}(\mathbf{h}_{sr}) = 1$, by the singular value decomposition (SVD) of $\mathbf{h}_{sr}\mathbf{h}_{sr}^H$, we have that $\mathbf{h}_{sr}\mathbf{h}_{sr}^H = \mathbf{U}_1\boldsymbol{\Sigma}_1\mathbf{U}_1^H$, $\boldsymbol{\Sigma}_1 = \text{diag}\left(||\mathbf{h}_i||^2, 0, \cdots, 0\right)$, $\mathbf{U}_1 = \left[\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \cdots, \mathbf{u}_{1,N}\right]$ where $\mathbf{U}_1$ notes the unitary matrix and $\mathbf{u}_{1,1} = \mathbf{h}_{sr}/||\mathbf{h}_{sr}||$, $\boldsymbol{\Sigma}_1$ the eigenvalue matrix, correspondingly. According to generalized Rayleigh entropy formula, we can get that

$$\mathbf{w}_{sr}^H\mathbf{h}_{sr}\mathbf{h}_{sr}^H\mathbf{w}_{sr} \leq \max\left\{\lambda\left(\mathbf{h}_{sr}\mathbf{h}_{sr}^H\right)\right\} = \max\{\lambda(\boldsymbol{\Sigma}_1)\} = \|\mathbf{h}_{sr}\|^2 \tag{10}$$

and the equal sign is applied in (10) when $\mathbf{w}_{sr} = \mathbf{u}_{1,1} = \mathbf{h}_{sr}/||\mathbf{h}_{sr}||$. So the expression of γ_{sr} and γ_{sd} can be written as $\gamma_{sr} = \bar{\gamma}_s|\mathbf{h}_{sr}|^2, \gamma_{sd} = \bar{\gamma}_s|h_{sd}|^2$, where $\bar{\gamma}_s = \frac{P_s}{\sigma_1^2}$ denotes the average SNR at S, and we already know that $|\mathbf{h}_{sr}|^2 = |h_1|^2 + |h_2|^2 + \cdots + |h_N|^2$. Meanwhile, the probability density functions (PDF) of $|h_{sd}|^2$ and $|h_i|^2 (i = 1, 2 \cdots N)$ are given by [6],

$$f_{|h_v|^2}(x) = \alpha_v \exp(-\beta_v x)\,{}_1F_1(m_v; 1; c_v x)\ (v \triangleq sd, 1, 2 \cdots N) \tag{11}$$

where $\alpha_v = (2b_v m_v/(2b_v m_v + \Omega_v))^{m_v}/2b_v$, $\beta_v = 1/2b_v$, $c_v = (\Omega_v/(2b_v m_v + \Omega_v))/2b_v$, $(v \triangleq sd, 1, \cdots N)$ with Ω_v being the average power of the LOS component, $2b_v$ the average power of the multipath component, m_v the Nakagami-m parameter. For analytical tractability, in the rest of this paper, we assume that $\Omega_{sd} = \Omega_1 = \cdots = \Omega_N$, $2b_{sd} = 2b_1 = \cdots = 2b_N$, $m_{sd} = m_1 = \cdots = m_N$, so vth can be dropped, meanwhile, we retain our focus in the case when the Nakagami-m parameter takes integer values, *i.e.* $m \in \mathbb{N}$. Hence, with the help of [7], ${}_1F_1(m; 1; cx)$ become

$${}_1F_1(m; 1; cx) = \exp(cx) \times \sum_{k=0}^{m-1} \frac{(-1)^k (1-m)_k (cx)^k}{(k!)^2} \tag{12}$$

where $(x)_n = x(x+1)\cdots(x+n-1)$. Then, by applying (12) to (11), the PDF of $|h_v|^2$ have the formulas as

$$f_{|h_v|^2}(x) = \sum_{k=0}^{m-1} \underbrace{\frac{(-1)^k (1-m)_k c^k}{(k!)^2}}_{\xi(k)} \times \alpha x^k \exp(-(\beta - c)x),\ (v \triangleq sd, 1, 2 \cdots N) \tag{13}$$

after some mathematical calculation, we have that [8]

$$f_{|\mathbf{h}_{sr}|^2}(x) = \sum_{k_1=0}^{m-1} \cdots \sum_{k_N=0}^{m-1} \Xi(N) x^{\Lambda-1} \exp(-(\beta - c)x) \tag{14}$$

where $\Lambda \triangleq \sum_{i=1}^{N} k_N + N$ and $\Xi(N) = \prod_{i=1}^{N} \xi(k_i)\alpha^N \prod_{j=1}^{N-1} B(\sum_{l=1}^{j} k_l + j, k_{j+1} + 1)$, so the PDFs of γ_{sd} and γ_{sr} can be written as

$$f_{\gamma_{sd}}(x) = \sum_{k=0}^{m-1} \frac{\xi(k)}{\overline{\gamma}_s^{k+1}} \alpha x^k \exp\left(\frac{-(\beta - c)x}{\overline{\gamma}_s}\right) \tag{15}$$

$$f_{\gamma_{sr}}(x) = \sum_{k_1=0}^{m_1-1} \cdots \sum_{k_N=0}^{m_1-1} \frac{\Xi(N) x^{\Lambda-1}}{\overline{\gamma}_s^{\Lambda}} \exp\left(-\left(\frac{\beta - c}{\overline{\gamma}_s}\right)x\right) \tag{16}$$

3.2 Terrestrial Channel

Since the R-D link is supposed to undergo correlated Rayleigh fading, which is superposition of L paths, the $\mathbf{h}_{rd}(N \times 1)$ can be modeled as [9]

$$\mathbf{h}_{rd}(t) = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} \rho_l(t) \mathbf{a}(\theta_l) \tag{17}$$

where $\theta_l \sim U\left(\bar{\theta}_l \pm \frac{\Delta\theta_l}{2}\right)$ and $\rho_l \sim \mathbb{CN}(0, \sigma^2)$denote the direction-of-arrival and the fading coefficient of the lth path signal, and $\mathbf{a}(\theta_l)$ is giving by

$$\mathbf{a}(\theta_l) = \left[1, \exp(j\kappa d_a \cos\theta_l), \ldots, \exp(j(N-1)\kappa d_a \cos\theta_l)\right]^T \tag{18}$$

with $\kappa = \frac{2\pi}{\lambda_{rd}}$ the wavenumber, λ_{rd} the carrier wavelength and d_a the relay's antenna space. Then, we apply MRT protocol to transmit the signal to D, and from the inference in last section, it can be deduced that $\mathbf{w}_{rd} = \frac{\mathbf{h}_{rd}}{\|\mathbf{h}_{rd}\|}$. In order to get the PDF of γ_{rd}, we first use the Kronecker model, so we can get that $\mathbf{h}_{rd} = \mathbf{R}^{\frac{1}{2}}\tilde{\mathbf{h}}_{rd}$, where $\tilde{\mathbf{h}}_{rd} = \left[\tilde{h}_{rd,1}, \ldots, \tilde{h}_{rd,N}\right]^T$ with $\tilde{h}_{rd,i} \sim \mathbb{CN}(0,1)$ are i.i.d random variables, and $\mathbf{R}$ is the covariance matrix of $\mathbf{h}_{rd}$

$$\mathbf{R} = \mathrm{E}\left[\mathbf{h}_{rd}(t)\mathbf{h}_{rd}^H(t)\right] = \frac{1}{L}\sum_{l=1}^{L} \mathrm{E}\left[|\rho_l(t)|^2\right] \mathbf{a}(\theta_l)\mathbf{a}^H(\theta_l) \tag{19}$$

Then, we can denote γ_{rd} from (6)

$$\gamma_{rd} = \bar{\gamma}_r \left|\mathbf{w}_{rd}^H \mathbf{h}_{rd}\right|^2 = \bar{\gamma}_r |\mathbf{h}_{rd}|^2 = \bar{\gamma}_r \left(\mathbf{R}^{\frac{1}{2}}\tilde{\mathbf{h}}_{rd}\right)^H \left(\mathbf{R}^{\frac{1}{2}}\tilde{\mathbf{h}}_{rd}\right) = \bar{\gamma}_r \tilde{\mathbf{h}}_{rd}^H \underbrace{\mathbf{R}^{\frac{H}{2}}\mathbf{R}^{\frac{1}{2}}}_{\mathbf{Q}} \tilde{\mathbf{h}}_{rd} \tag{20}$$

with $\gamma_r = \frac{P_s}{\sigma_2^2}$ presents the average transmit SNR at R. By the SVD of Q, we have that

$$\mathbf{Q} = \mathbf{R}^{\frac{H}{2}}\mathbf{R}^{\frac{1}{2}} = \mathbf{U}_2\mathbf{\Sigma}_2\mathbf{U}_2^H,\ \mathbf{U}_2 = \left[\mathbf{u}_{2,1}, \ldots, \mathbf{u}_{2,N}\right],\ \mathbf{\Sigma}_2 = \mathrm{diag}\left(\lambda_{2,1}, \ldots, \lambda_{2,N}\right) \tag{21}$$

where $\mathbf{U}_2$ denotes the unitary matrix, $\mathbf{\Sigma}_2$ the eigenvalue matrix, correspondingly, with $\lambda_{2,j}(j = 1, \ldots, N)$ the eigenvalues which is arranged in decreasing order. After some mathematic calculation, γ_{rd} can be denoted from (20), (21)

$$\gamma_{rd} = \bar{\gamma}_r \sum_{i=1}^{N} \lambda_{2,i} \left| \tilde{h}_{rd,i} \right|^2 \tag{22}$$

With the help of Laplace transformation, the PDF of γ_{rd} can be written as

$$f_{\gamma_{rd}}(x) = \sum_{p=1}^{t} \sum_{q=1}^{v_p} \frac{a_{p,q}}{\Gamma(q)\left(\lambda_{2,p}\bar{\gamma}_r\right)^q} x^{q-1} e^{-\frac{x}{\lambda_{2,p}\bar{\gamma}_r}} \tag{23}$$

where $\Gamma(q)$ is the Gamma function, t is the number of distinct nonzero eigenvalues and v_p denote the repeated times of $\lambda_{1,p}$, besides,

$$a_{p,q} = \frac{1}{(v_p - q)!\lambda_{2,p}^{v_p - l}} \frac{\partial^{v_p - l}}{\partial s^{v_p - l}} \left[\prod_{n=1, n\neq p}^{t_i} \frac{1}{1 + s\lambda_{2,n}} \right] \Bigg|_{s=-\lambda_{2,p}^{-1}} \tag{24}$$

In this paper, $v_p = 1$ invariably, so the PDF of γ_{rd} can be denoted as

$$f_{\gamma_{rd}}(x) = \sum_{p=1}^{t} \frac{a_p}{\lambda_{2,p}\bar{\gamma}_r} \exp\left(-\frac{x}{\lambda_{2,p}\bar{\gamma}_r} \right) \tag{25}$$

4 Outage Probability Analysis

The Outage Probability is defined as the probability that the SNR falls below a certain threshold value i.e.γ_{th}, such that

$$P_{\gamma_{th}}^{out} \triangleq \Pr\{\gamma < \gamma_{th}\} = (1 - (1 - F_{sr}(\gamma_{th})) \times (1 - F_{rd}(\gamma_{th}))) \times F_{sd}(\gamma_{th}) \tag{26}$$

and $\mathrm{F}_{\gamma_i}(x)(i \triangleq sr, sd, rd)$ is the cumulative distribution function (CDF) of γ_i, with the help of [10] and (15), we can found that

$$F_{\gamma_{sd}}(\gamma_{th}) = \sum_{k=0}^{m-1} \frac{k!\xi(k)\alpha}{(\beta - c)^{k+1}} \left(1 - \sum_{t=0}^{k} \frac{(\beta - c)^t \gamma_{th}^t}{t!\bar{\gamma}_s^t} \exp\left(\frac{-(\beta - c)\gamma_{th}}{\bar{\gamma}_s} \right) \right) \tag{27}$$

Meanwhile, following (16), (25), and [10], the CDFs of γ_{sr} and γ_{rd} can be written as

$$F_{\gamma_{sr}}(x) = 1 - \sum_{k_1=0}^{m-1} \cdots \sum_{k_N=0}^{m-1} \sum_{r=0}^{\Lambda-1} \frac{(\Lambda - 1)!\Xi(N)x^r}{r!\bar{\gamma}_s^r(\beta - c)^{\Lambda - r}} \exp\left(-\left(\frac{\beta - c}{\bar{\gamma}_s} \right) x \right) \tag{28}$$

$$F_{\gamma_{rd}}(x) = 1 - \sum_{p=1}^{t} a_p \exp\left(-\frac{x}{\lambda_{2,p}\bar{\gamma}_r}\right) \tag{29}$$

Then, put (27)–(29) into (26), and after some algebraic manipulations, we obtain the closed-form expression of the OP in this system

$$P_{ins}^{out} = \left[1 - \sum_{k_1=0}^{m-1} \cdots \sum_{k_N=0}^{m-1} \sum_{r=0}^{\Lambda-1} \sum_{p=1}^{t} \frac{(\Lambda-1)!\Xi(N) a_p \gamma_{th}^r}{r!\bar{\gamma}_s^r (\beta-c)^{\Lambda-r}} \exp\left(-\left(\frac{\beta-c}{\bar{\gamma}_s} + \frac{1}{\lambda_{2,p}\bar{\gamma}_r}\right)\gamma_{th}\right)\right] \times \left[\sum_{k=0}^{m-1} \frac{k!\xi(k)\alpha}{(\beta-c)^{k+1}}\left(1 - \sum_{t=0}^{k} \frac{(\beta-c)^t \gamma_{th}^t}{t!\bar{\gamma}_s^t} \exp\left(\frac{-(\beta-c)\gamma_{th}}{\bar{\gamma}_s}\right)\right)\right] \tag{30}$$

5 Numerical Results

This section conducts numerical simulation to demonstrate the validity of the theoretical expressions and show the impact of multi-antenna relay of the cooperative network. The simulation results are obtained with 10^7 channel realizations, and the analytical curves obtained by (34), and we assume $\bar{\gamma} = \bar{\gamma}_r = 2\bar{\gamma}_s$ for all the links.

Figure 2 illustrates the outage probability of cooperative network in which the satellite links undergoe average shadowing ($b = 0.126, m = 10, \Omega = 0.835$) [6]. Besides, in the R-D link, we assume that $L = 10$, $\Delta\theta = 5°$. For this case, we can clearly observe that the analytical outage probability expressions are in excellent agreement with the simulation results. As we expect, the OP of cooperative network decreases with the increasing of relay's antennas number N.

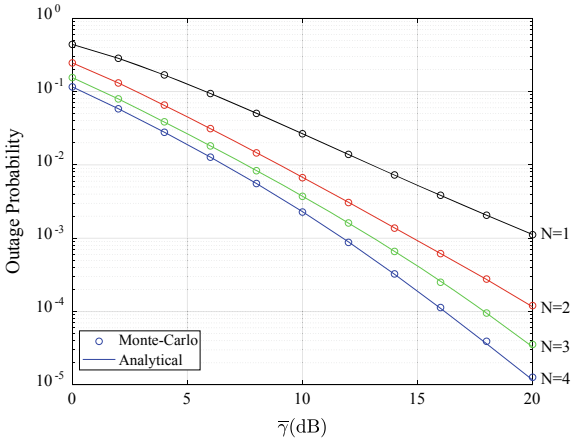


Fig. 2. OP of the system

6 Conclusion

In this paper, we investigated the performance of a HSTCN where the terrestrial relay equipped with N antennas, where the terrestrial channel undergoes the correlated Rayleigh fading. The expressions of OP have been derived, and the validity of the expression has been proved by comparison with Monto Carlo simulations, and it suggests the superiority of the multi-antennas relay in the HSTCN.

References

1. Etcharte, V., Vale, E.: The integration of the satellite communications with the terrestrial mobile network (UMTS). IEEE Lat. Am. Trans. **10**(1), 1175–1179 (2012)
2. An, K., Lin, M., Liang, T.: On the performance of multiuser hybrid satellite-terrestrial relay networks with opportunistic scheduling. IEEE Commun. Lett. **19**(10), 1722–1725 (2015)
3. Bhatnagar, M.R., Arti, M.K.: Performance analysis of AF based hybrid satellite-terrestrial cooperative network over generalized fading channels. IEEE Commun. Lett. **17**(10), 1912–1915 (2013)
4. Yuksel, M., Erkip, E.: Diversity in relaying protocols with amplify and forward. In: Global Telecommunications Conference, GLOBECOM 2003, vol. 4, pp. 2025–2029. IEEE (2003)
5. Ruan, Y., Li, Y., Zhang, R., Zhang, H.: Performance analysis of hybrid satellite-terrestrial cooperative networks with distributed alamouti code. In: 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring), pp. 1–5 (2016)
6. Abdi, A., Lau, W.C., Alouini, M.S., Kaveh, M.: A new simple model for land mobile satellite channels: first- and second-order statistics. IEEE Trans. Wirele. Commun. **2**(3), 519–528 (2003)
7. Wolfram Research: Mathematica Edition, 8.0 edn. (2010)
8. Miridakis, N.I., Vergados, D.D., Michalas, A.: Dual-hop communication over a satellite relay and shadowed Rician channels. IEEE Trans. Vehicular Technol. **64**(9), 4031–4040 (2015)
9. Lin, M., Ouyang, J., Zhu, W.P.: Joint beamforming and power control for device-to-device communications underlaying cellular networks. IEEE J. Sel. Areas Commun. **34**(1), 138–150 (2016)
10. Gradshteyn, I.S., Ryzhik, I.M.: Table of Integrals, Series, and Products. Academic, San Diego (2007)

A Cross-Layer Image Transmission Method for Deep-Space Exploration

Dongqing Li, Shaohua Wu(✉), Jian Jiao, and Qinyu Zhang

ShenZhen Graduate School, Harbin Institute of Technology, Shenzhen, China
lidongqing@stu.hit.edu.cn
{hitwush,jiaojian,zqy}@hit.edu.cn

Abstract. High efficiency and reliable image transmission is critical for deep-space communications. In the traditional deep-space image transmission system design, the work is mainly focused on the compression efficiency and the coding gain, respectively, while neglecting the implementation complexity. To improve the transmission efficiency of images in deep-space communication with a long delay, while satisfying the affordable implementation complexity, cross-layer design is critical. In this paper, we propose a novel high-efficiency cross-layer image transmission method for deep-space exploration. The proposed method is designed based on compressed sensing (CS) in the application layer, the Spinal codes in the physical layer, and the Licklider transmission protocol (LTP) in the transport layer. By jointly optimizing across the application, transport, and physical layers, we consider the time-varying deep-space channel and proposed a CS-error-tolerant rate-adaptive strategy based on the robustness of CS to error. Extensive simulations are carried out for performance evaluation. Results show that the proposed cross-layer image transmission method with the CS-error-tolerant rate-adaptive strategy can significantly improve the performance of transmission efficiency.

Keywords: Deep-space communication · Image transmission · Cross-layer design · CS-error-tolerant · Rate-adaptive strategy

1 Introduction

One of the core tasks of deep-space communication is to send immense amounts of image data, such as the camera-captured images of the surface of a remote planet, back to the Earth ground station. However, deep-space communication is characterized by features of large signal attenuation, time-varying propagation delay, high error rate, and limited node resources. When considering the critical conditions of deep-space links [1], one needs to pay particular attention to improving the efficiency of the deep-space image transmission system. Motivated by the idea of cross-layer design for terrestrial networks, in this paper, we aim to propose an efficient cross-layer deep-space image transmission method

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 257–264, 2019.
https://doi.org/10.1007/978-981-13-6264-4_31

in the DTN framework [2]. The proposed method is designed based on cross optimization over the application, transport, and physical layers.

In the application layer for image compression, we adopt compressed sensing (CS) technique [3], known for its advantages including the low-encoding complexity and the scalable compression ability to realize low-encoding complexity yet high compression efficiency. In the physical layer for error protection, we adopt the newly invented rateless Spinal codes [4], which can achieve the capacities of both the Additive White Gaussian Noise (AWGN) channel and the Binary Symmetric Channel (BSC) to realize affordable coding complexity. In the transport layer, the newly developed Licklider transmission protocol (LTP) of DTN is designed to operate over point-to-point, long-haul, frequent interruption links [5], and it could tolerate these characteristics with no reliance on the stability of the communication round trip time (RTT) [6]. Therefore, LTP is naturally used for the transmission control in our proposed method.

In a nutshell, the proposed method incorporates the CS for image compression and the Spinal codes for error protection into the DTN protocol stack, to jointly work with the LTP. On the basis of this framework, we propose the cross-layer transmission method and form a CS-error-tolerant rate-adaptive transmission strategy aiming for throughput maximization as an optimization problem. Extensive simulations are carried out to show that the proposed method can significantly improve the efficiency of deep-space image transmission and outperform the other methods.

The main contributions of the paper are summarized as follows:

- We propose a cross-layer transmission method for highly efficient deep-space image transmission. The CS image compression and the Spinal codes are incorporated in the framework to jointly work with the LTP.
- We design a CS-error-tolerant rate-adaptive transmission strategy for the image transmission system to dynamically match the time-varying deep-space channels.
- We propose a model to formulate the optimal transmission strategy that can maximize the downlink image transmission throughput.

The rest of this paper is organized as follows. In Sect. 2, we present the details of cross-layer deep-space image transmission method and the CS-error-tolerant transmission strategy, and the optimal transmission strategy designs are formulated. In Sect. 3, the case study on the Earth–Mars communication scenario is presented. The simulation results are given in Sect. 4. The conclusions are drawn in Sect. 5.

2 The Proposed Cross-Layer Image Transmission Method

2.1 The Cross-Layer Transmission Framework

As illustrated in Fig. 1, the downlink image transmission in deep-space communications typically involves the source node, the relay nodes, and the destination node. In the process of image transmission system, the space probe node performs

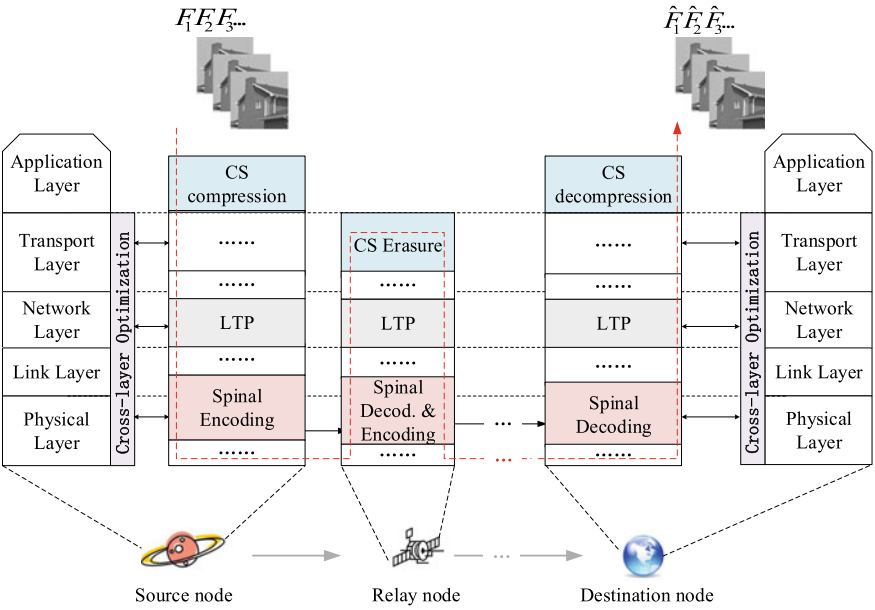


Fig. 1. The framework of cross-layer deep-space image transmission system.

the CS image compression. After a sequence of compressed values are generated by the CS image compression modules, they are sent to the lower layers of the DTN stack. Then the bitstreams are encoded by the Spinal codes in the physical layer, the Spinal symbols are sent to the space relay node. On receiving the Spinal symbols, the space relay node conducts decoding sequentially, and takes CRC to check on each decoded block. Meanwhile, the space relay node checks whether the packet error rate (PER) can satisfy the requirement of a certain CS decompression PSNR. If it is satisfied, the data will be re-encapsulated and sent to the ground station node, otherwise the space relay node will send feedback to the space probe node and request additional Spinal symbols. Finally, the ground station node completes the final decoding, decapsulation, and reconstruction to obtain the deep-space images. By cross-layer design, the CS compression in the application layer, the Spinal channel coding in the physical layer, and the LTP transmission control will be jointly optimized to achieve a gain in transmission efficiency.

2.2 CS-Error-Tolerant Rate-Adaptive Transmission Strategy

To fully explore the transmission effciency of time-varying deep-space channel, the transmission strategy should not aim only at fixed channel state. Therefore, we propose a rate-adaptive transmission strategy to realize dynamic matching with the time-varying deep-space channels.

Due to the time-varying and extremely low SNR of deep-space communication links, the encoded data are difficult to transmit reliably, i.e., the packet

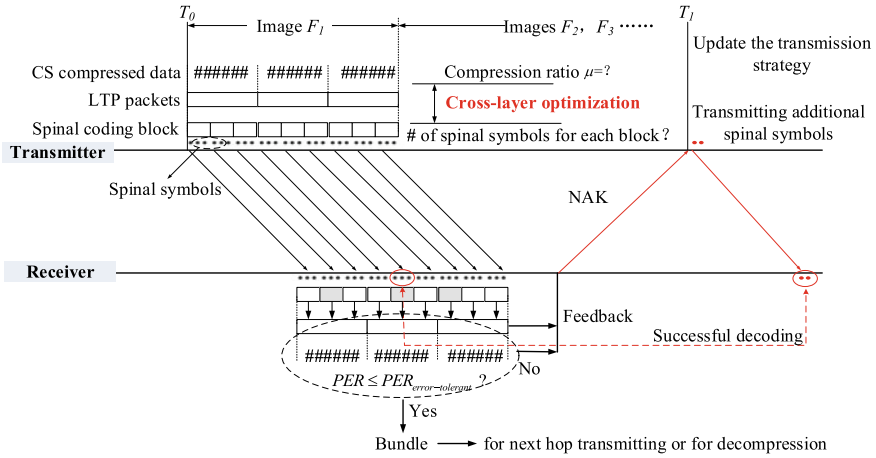


Fig. 2. Illustration of the cross-layer image transmission process based on CS-error-tolerant strategy.

error rate is still high after physical layer decoding. In order to ensure the data passed to the application are correct, the conventional transmission strategy simply discards the erroneous data blocks which cannot pass the CRC check. Although this way guarantees that the data passed to the application layer are perfectly correct, it also imposes all the burden of the error protection on the bottom layer, and indirectly increases the number of feedbacks. To improve the overall performance of the image transmission system, we introduce the CS-error-tolerant based rate-adaptive strategy to the cross-layer image transmission method. The CS-error-tolerant strategy allows the erroneous blocks to be passed to the application layer for CS reconstruction together with the correctly decoded data packets. In other words, we expect to take advantage of the error-tolerant ability of compressed sensing to further improve the image transmission performance.

2.3 Cross-Layer Optimization Transmission Strategy

Figure 2 illustrates the detailed process of the proposed cross-layer transmission method based on the CS-error-tolerant strategy. The original image is first submitted to the CS compression module with a compression ratio μ given by Eq. (1):

$$\mu = (M \cdot Q)/(N \cdot P), \tag{1}$$

where N is the number of pixels in the original image, P is quantization precision of each pixel, M is the number of compression values for each image, and Q is the quantization precision of each compression value.

When an image has been encapsulated by the LTP protocol, the number of LTP segments can be determined from Eq. (2):

$$N_{\text{Seg}} = \mu \cdot N \cdot P/(l_{\text{Seg}} - l_{\text{Hdr}}), \tag{2}$$

where l_{Seg} is the length per segment packaged by the LTP protocol and l_{Hdr} is the length of the packet header. Moreover, the number of Spinal coded blocks in each segment is n_{CB}, which can be expressed according to Eq. (3):

$$n_{\mathrm{CB}} = (l_{\mathrm{Seg}} - l_{\mathrm{Hdr}})/l_{\mathrm{CB}}, \tag{3}$$

where l_{CB} is the length of one Spinal coded block. Let the total number of Spinal coded blocks corresponding to one image be denoted by N_{CB}, given by Eq. (4):

$$N_{\mathrm{CB}} = N_{\mathrm{Seg}} \cdot n_{\mathrm{CB}}. \tag{4}$$

Based on the fact that the packet-level deep-space channel can be modeled as finite-state Markov chains [7], we divide the channel conditions into T Markov-states, and the t^{th} channel state can be expressed as $C_t = j (j \in \{1, 2, ..., T\})$. When all the decoded blocks in a segment pass the CRC check, the segment can be regarded as successfully received, and its probability is $\mathrm{CDF}(n_{\mathrm{Spinal}}|C_t = j)^{n_{\mathrm{CB}}}$ $(j \in \{1, 2, ..., T\})$. Therefore, the packet error rate of the received data can be expressed as $1 - \mathrm{CDF}(n_{\mathrm{Spinal}}|C_t = j)^{n_{\mathrm{CB}}}$. The channel transition probability matrix is P. When the sender opens the next session, the packet error rate of the LTP data segments can be expressed as Eq. (5):

$$PER = 1 - [\mathrm{CDF}(n_{\mathrm{Spinal}}|C_t = 1)^{n_{\mathrm{CB}}}, \cdots \mathrm{CDF}(n_{\mathrm{Spinal}}|C_t = T)^{n_{\mathrm{CB}}}] \cdot P. \tag{5}$$

In order to ensure that the quality of the reconstructed image can reach a certain PSNR, the constraint is given by Eq. (6):

$$PER \leq PER_{\mathrm{error-tolerant}}. \tag{6}$$

where $PER_{\mathrm{error-tolerant}}$ is the maximum affordable packet error rate for a certain reconstruction PSNR when the current compression ratio is μ. The object to be minimized is the total number of Spinal symbols for each image, i.e., given by Eq. (7):

$$N_{\mathrm{Spinal}} = N_{\mathrm{CB}} \cdot n_{\mathrm{Spinal}}. \tag{7}$$

The parameters to be solved for the minimization are the CS compression ratio, denoted by μ, and the number of Spinal codes per coded block, denoted by n_{Spinal}. The optimal values of μ and n_{Spinal} can be solved using Eq. (8):

$$\begin{aligned} &\mu, n_{\mathrm{Spinal}} = \underset{\mu, n_{\mathrm{Spinal}}}{\arg \min} \{N_{\mathrm{Spinal}}\} \\ &\text{s.t} \quad \text{equation(6)}. \end{aligned} \tag{8}$$

3 Case Study on the Optimal Transmission Strategy

As shown in Fig. 3, we select the Gilbert-Elliot model which is a two-state Markov model to simulate the whole transmission channel approximately. In [7] the authors used meteorological statistics, and they concluded that 20 K is

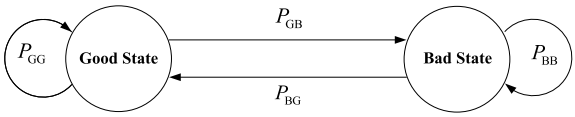


Fig. 3. Gillbert-Elliot model.

Table 1. Result of the optimization problem for CS-error-tolerant transmission strategy in Gilbert-Elliot channel model

Parameter	Good state	Bad state
μ	0.61	0.68
n_{Spinal}	352	768

the threshold that divides the channel states. In the GE model, when the channel is in the Good state, the corresponding bit error rate is e_{Good}. And when burst errors occur, the corresponding bit error rate is e_{Bad}. They gave the corresponding recommended bit error rate values as follows: e_{Good} can be set as 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5} and e_{Bad} can be set as 10^{-4}, 10^{-3}, the values of channel state transition probabilities are also given by Eq. (9):

$$P = \begin{bmatrix} P_{\text{GG}} & P_{\text{GB}} \\ P_{\text{BG}} & P_{\text{BB}} \end{bmatrix}, \tag{9}$$

where $P_{\text{GG}} = 0.9773$, $P_{\text{GB}} = 0.0227$, $P_{\text{BB}} = 0.8333$, $P_{\text{BG}} = 0.1667$.

We further consider parameter settings about the deep-space image as follows: the original image is sized by 512*512, 8 bits per pixel, and the reconstruction algorithm is BM3D-AMP, the quantization precision is 5 bits. And the minimum CS image reconstruction PSNR is set to 30 dB. Moreover, l_{Seg} is 1472 bytes, l_{HDR} is 16 bytes, and l_{CB} is 32 bytes. The parameters of the Spinal codes are $n = 256$, $k = 4$, $d = 2$, $B = 32$. Based on the parameter values of the GE model and the settings of image transmission parameters, the optimal transmission strategy in each channel state can be solved using non-linear optimization tools, such as the Matlab 'fmincon' toolbox. The result is shown in Table 1.

4 Results and Discussion

We conducted a series of simulation experiments and obtained the throughput results for different transmission methods. The simulated scenario is the Earth–Mars image transmission. The parameter settings are the same as the case study settings in Sect. 3.

Five different transmission methods, including the proposed optimal cross-layer methods are simulated for comparisons. They are listed as follows: (1) The proposed CS-error-tolerant optimal cross-layer method. This method allows the error packets to be passed to the application layer in order to further improve the

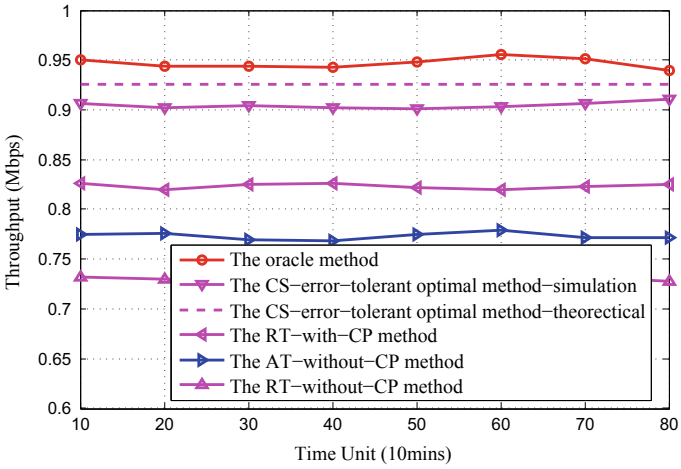


Fig. 4. Performance results comparison for five different transmission methods.

system efficiency. (2) The retransmission on the NAK method without channel state prediction. This method is actually the one commonly used in existing deep-space communications. It is different from the proposed method in two aspects. First, on receiving NAK the transmitter does a retransmission instead of additionally sending data. The second difference is that the transmitter does not predict the future channel state. (3) The retransmission on the NAK method with channel state prediction. (4) The additional-transmission on NAK without channel state prediction. (5) The transmission method of an oracle transmitter. By 'oracle', the transmitter is assumed to know everything about the channel and the receiver, so it can precisely adjust the optimal transmission strategy, and the receiver can ideally decode all the data successfully. This method is the idealized transmission method. It is used as the performance upper bound in this work.

Figure 4 shows the comparison among the optimal method and the other existing methods. It can be seen that the throughput of the optimal method is very close to the oracle method. This indicates that through the cross-layer optimization, we could realize the high-efficient image transmission. Therefore, the result validates the effectiveness of the proposed optimal method. When comparing the optimal method and the RT-without-CP method which is commonly used in existing deep-space communications, the throughput of the optimal method-simulation increases by 20% with respect to the RT-without-CP method, implying that the proposed optimal method can obtain a significant gain of transmission efficiency.

5 Conclusion

In this paper, we propose a novel cross-layer deep-space image transmission method where the CS image compression and the Spinal codes are incorporated in the system to jointly work with the LTP under the DTN protocol stack. Based on the proposed image transmission system, we design a rate-adaptive transmission strategy to match the time-varying deep-space channels and provide cross-layer optimization model of the transmission strategy. In particular, the obtained results reveal that the cross-layer design can significantly improve the performance of transmission efficiency, comparing with the other methods.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant 61371102, 61771158, and in part by the Shenzhen Municipal Science and Technology Plan under Grant JCYJ20170811160142808, JCYJ20170811154309920.

References

1. Huntress, W., Stetson, D., Farquhar, R., Zimmerman, J., Clark, B., Neil, W.O., Bourke, R., Foing, B.: The next steps in exploring deep space a cosmic study by the IAA. Acta Astronaut. **58**(6–7), 304–377 (2006)
2. Nie, Y.L., Peng, F.S., Zhang, G.X., Yong-Qiang, L.I., Xie, Z.D., Jing, H.U.: Research on application of DTN protocol architectures in deep space communication. In: Radio Communications Technology (2016)
3. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theory **52**(2), 489–509 (2006)
4. Perry, J., Iannucci, P.A., Fleming, K.E., Balakrishnan, H., Shah, D.: Spinal codes. In: Proceedings of the ACM SIGCOMM 2012 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, pp. 49–60. ACM (2012)
5. Wang, R., Parikh, P., Bhavanthula, R., Hou, J., Zhou, L., Taleb, T.: Licklider transmission protocol (LTP)-based DTN for long-delay cislunar communications. In: Global Telecommunications Conference (GLOBECOM, 2010), pp. 1–5. IEEE (2010)
6. Yang, Z., Wang, R., Yu, Q., Sun, X., De Sanctis, M., Zhang, Q., Hu, J., Zhao, K.: Analytical characterization of licklider transmission protocol (LTP) in cislunar communications. IEEE Trans. Aerosp. Electron. Syst. **50**(3), 2019–2031 (2014)
7. Gao, J.L., et al.: CFDP performance over weather-dependent Ka-band channel (2006)

Research of Turbo Codes in Unmanned Aerial Vehicle Data Link System

Weizheng An[1(✉)], Ruina Zhao[2], Linghui Zhang[1], Lei Guo[1], and Shao Li[1]

[1] China North Vehicle Research Institute, Beijing, China
anweizh@163.com, {zhlh,guolei,lishao}@noveri.com.cn
[2] Beijing North Vehicle Group Corporation, Beijing, China
zhaoruina2011@163.com.cn

Abstract. In this paper, we will discuss the use of Turbo codes in unmanned aerial vehicle (UAV) data link system. High-quality communication is one of the key technologies to ensure the effectiveness of UAV data link. According to the characteristics and requirements of UAV's, the Turbo decoding model is studied, and Matlab simulation is carried out for various parameter conditions. At the same time, FPGA is used to achieve the Turbo encoder and decoder, and complete the verification analysis and testing.

Keywords: UAV data link · Turbo codes · FPGA · Signal to noise ratio · Turbo decoding

1 Introduction

In recent years, with the UAV in military, technology, environmental and other areas of the extensive use of UAV monitoring and control technology development usher in a golden age. The performance of the UAV data link system is constrained by the radio channel. According to the wireless communication environment, the UAV data link system is a frequency selective Rayleigh channel or a Rice channel. Therefore, the decoding algorithm under the fading channel is the key to applying efficient channel coding to UAV data link system [1].

In the modern wireless communication system, some form of error control coding is used to improve the reliability of the communication system. UAV monitoring and control data link is no exception. Turbo codes have excellent decoding performance close to the Shannon limit, which has become a hot topic in the coding field. But some of the characteristics of Turbo code limits its application in some communication systems.

In this paper, we analyzed the performance of Turbo codes, and design Turbo code based on FPGA. We make Turbo code design more simple and better meet the UAV data link for high-speed, high-reliability communication requirements.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 265–273, 2019.
https://doi.org/10.1007/978-981-13-6264-4_32

2 UAV Channel Model

In the UAV channel propagation environment, the received signal is usually a composite signal of the multipath signal due to the phenomena such as reflection, diffraction and scattering caused by various obstacles between the UAV and the control station. The random fluctuation of the phase, amplitude and arrival time of the multipath signal will cause the fluctuation of the envelope of the received signal [2]. The actual measurement data analysis shows that the fast fading pattern of this received signal envelope amplitude follows the Rician distribution. So the UAV channel can be expressed by the Rician fading channel model. The output signal r(t) is expressed as

$$r(t) = (k + a)s(t) + n(t) \tag{1}$$

where s(t) is the input signal, R(t) is the output signal, n(t) is additive Gauss noise, K is the main fading factor, "a" is the other fading factor.

3 Turbo Code Coding and Decoding Principle

In 1993, C. Berrou, A. Glavieux and P. Thitima-jshima put forward the Turbo code, in the randomness of the encoding and decoding conditions to obtain nearly the theoretical limit of Shannon decoding performance [3]. Turbo code encoder through the interleaver to two recursive systems convolutional code parallel cascade, the decoder in the two component code decoder between the iterative decoding. The fundamental reason for the superior performance of the Turbo code is the use of iterative decoding to improve the decoding performance by exchanging soft information between component decoders. The iterative algorithm is used to make the decoding performance as close as possible to the maximum likelihood algorithm [4].

3.1 Turbo Coding Principle

Turbo codes use interleaver to achieve the idea of random coding. It combines the short code effectively into a long code. Its structure is shown in Fig. 1.

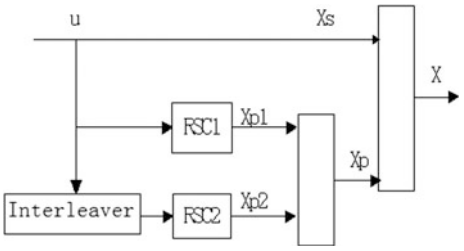


Fig. 1. Turbo coding principle

The encoder consists of two recursive system convolutional encoder (RSC1 and RSC2) parallel concatenation, the information sequence before reaching the RSC2 to be interleaved, and then through the deletion and multiplexing, resulting in different code rates.

3.2 Turbo Decoding Principle

The decoder uses the iterative decoding scheme, which is complementary to the parallel cascaded coding scheme. It can be improved obviously Decoding performance.

The most common decoding algorithms for Turbo codes are SOVA, max-log-MAP, log-MAP and MAP.

The MAP algorithm is the most complex in the four algorithms, and the Log-MAP algorithm greatly reduces the complexity of the algorithm because it transforms a large number of multiplication operations into additions. The Max-Log-MAP algorithm further reduces the complexity of the algorithm, but the performance is also reduced. The SOVA algorithm is the same as the basic method of the Max-Log-MAP algorithm. When v = 2, the SOVA algorithm is about Max-Log-MAP algorithm is twice, and when v = 4, Max-Log-MAP algorithm is almost twice the calculation of SOVA [5].

4 Research on Turbo Code Performance

The performance of the iterative decoding algorithm of Turbo codes is lack of effective theoretical explanation, and the application in UAV data link is lack of necessary research. In order to design the encoder and decoder which meets the characteristics of UAV data link, the most effective method is to use computer simulation.

4.1 Influence of Interleaving Length on Performance

Turbo codes have excellent performance because of the introduction of an interleaver in their codecs. Interleaver is an important part of the Turbo code system, making the Turbo code has an approximate random performance. The essence of the interleaver is to reset the position of each element in the sequence, so as to get the interleaving sequence.

The interleaving length was set to 128, 256, 512, 1024, 2048, coding efficiency is 1/3, the generation matrix (37, 21), decoder using Max-Log-MAP algorithm, the number of iterations for the 8. The simulation results are shown in Fig. 2.

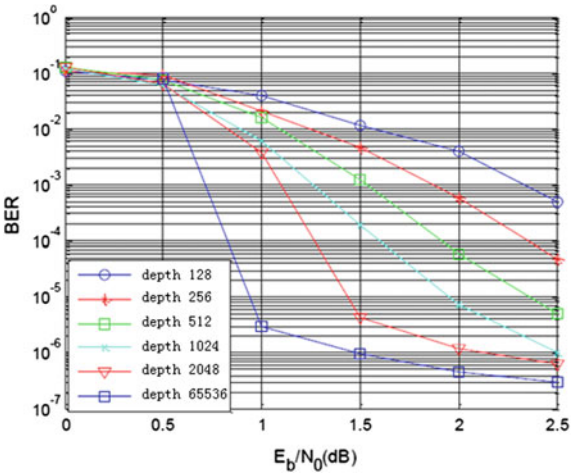


Fig. 2. Interleaving length performance impact

It can be seen from Fig. 2 that the length of the interleaver determines the key factors of the performance of the Turbo code when 10-6 ≤ BER ≤ 10-2. However, when the Eb/N0 value is small, the effect of increasing the interleaving length on the bit error rate performance is poor.

4.2 Influence of Number of Iterations on Performance

The performance of Turbo codes is directly affected by the number of decoding iterations.

The interleaving depth is 1024, coding efficiency is 1/3, the generation matrix (37, 21), decoder using Max-Log-MAP algorithm, the number of iterations is 1, 2, 3, 4, 5, 6, 7, 8. The simulation results are shown in Fig. 3.

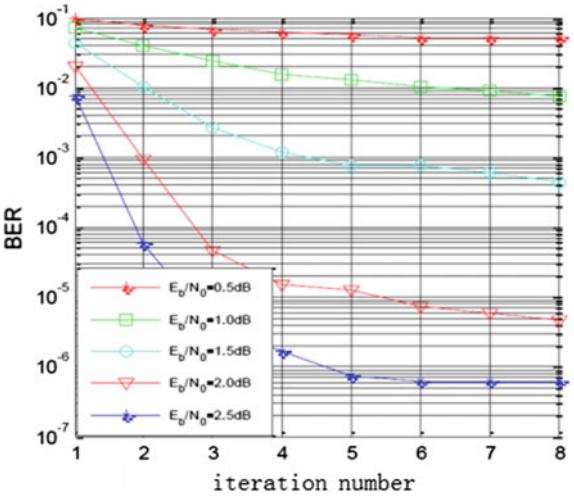


Fig. 3. Scheme of the high level with high resistance test circuit

It can be seen from Fig. 3 that as the number of iterations increases, the bit error rate decreases and tends to converge. With the increase of Eb/N0, the number of iterations on the bit error rate performance is more and more obvious. The initial iteration has a significant effect on the coding gain, and the value of the bit error rate tends to stabilize after the fourth and fifth iterations. When BER $\geq$ 10-2 the effect of iteration on BER is very small, and when 10-6 $\leq$ BER $\leq$ 10-2, the bit error rate Of the increase and decline, but after a certain number of iterations to become gentle.

5 FPGA Design of Turbo Codes

We use Xilinx FPGA chip for Turbo code design based on UAV data link. The decoding algorithm of Turbo codes has many disadvantages such as high decoding complexity and large delay. Therefore, the focus of the research is to implement high performance low complexity and delay Turbo decoder. FPGA has the advantages of fast computing speed, repeatable programming, powerful function, short circuit design cycle and low cost.

5.1 FPGA Implementation of Encoder

Turbo encoder is relatively simple. It mainly includes RSC component coder, interleaver, delete and reorder unit (Fig. 4).

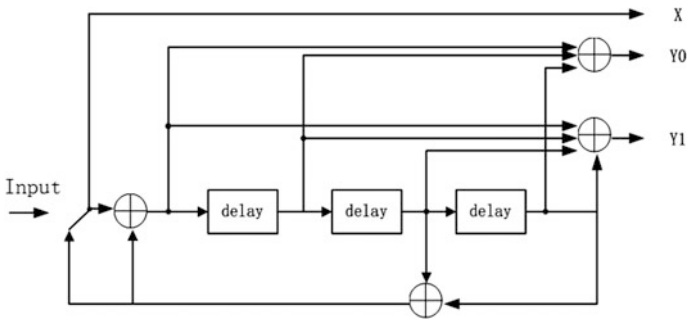


Fig. 4. RSC component coder structure

The interleaver is an important part of the performance and delay of Turbo codes. We construct an address generator to generate an interleaved address. While this solution needs to be wasted for several cycles in the implementation process, the operation reduces the need for storage resources.

Turbo code encoder simulation results shown in Fig. 5.

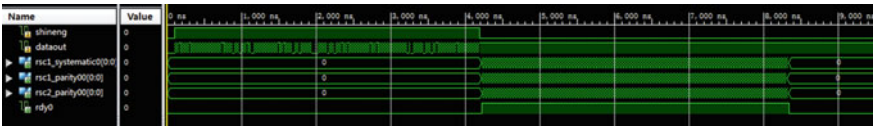


Fig. 5. RSC component coder structure

5.2 Decoder FPGA Implementation

The decoder mainly comprises a component decoder, an interleaved and deinterleaver and a plurality of memories. Wherein the interleaving and deinterleaver are implemented in the same way as the encoding section, the interleaving process is written in the order of the address of the interleaved pattern, and the deinterleaving process is written in the order of the interleaved pattern address.

Through theoretical calculation and simulation, the performance of the Log-MAP algorithm is better than the SOVA algorithm; Max-Log-MAP algorithm is further simplified to the Log-MAP algorithm, a decrease in performance, but still better than the SOVA algorithm. For the computational complexity, the computational complexity of Log-MAP algorithm is the largest, followed by Max-Log-MAP, and the SOVA is the smallest [6]. Although the computational complexity of the SOVA algorithm is low, there is a big drawback: the bit error rate cannot be decreased immediately after reaching a certain limit. Although the MAP algorithm has the same problem, it can meet the requirement of bit error rate. We choose the Max-Log-MAP decoding algorithm [7].

In the AWGN channel model, Max-Log-MAP algorithm is calculated as follows:

$$M_k(e) = \begin{cases} \Lambda_a(u_k) - \max(0, \Lambda_a(u_k)) + \frac{1}{2} * \frac{4\sqrt{E_s}}{N_0} y_k^s + \frac{1}{2} * \frac{4\sqrt{E_s}}{N_0} y_k^p \left(2c_k^p - 1\right), u_k = 1 \\ -\max(0, \Lambda_a(u_k)) - \frac{1}{2} * \frac{4\sqrt{E_s}}{N_0} y_k^s + \frac{1}{2} * \frac{4\sqrt{E_s}}{N_0} y_k^p \left(2c_k^p - 1\right), u_k = 0 \end{cases} \tag{2}$$

The forward, reverse and logarithmic likelihood ratios are expressed as follows:

$$A_k(s) = \max_{e, S^S(E)} \left[\mathrm{A_{k-1}S_k^E(s) + M_k(E)} \right], K = 1, 2, \ldots, \mathrm{N} - 1 \tag{3}$$

$$\mathrm{B_k(s)} \max_{\mathrm{e,S^S(E)}} \left[\mathrm{B_{k-1} | S_{k+1}^E(s) + M_{k+1}(E)} \right], \mathrm{K} = 1, 2, \ldots, \mathrm{N} - 1 \tag{4}$$

$$\begin{aligned} \Lambda_{\mathrm{k}}(\mathrm{u}, \mathrm{o}) &= \Lambda_{\mathrm{a}}(\mathrm{u_k}) + \Lambda_{\mathrm{k}}(\mathrm{c^s}, \mathrm{I}) \\ &+ \max_{\mathrm{e,u(e=1)}} \left[\frac{1}{2} \Lambda_{\mathrm{k}}(\mathrm{c^s}, \mathrm{I}) \left(2\mathrm{c_k^p} - 1\right) + \mathrm{A_{k-1}}\left(\mathrm{S_k^S(e)}\right) + \mathrm{B_k}(\mathrm{S_k^K(e)}) \right] \\ &- \max_{\mathrm{e,u(e=1)}} \left[\frac{1}{2} \Lambda_{\mathrm{k}}(\mathrm{c^s}, \mathrm{I}) \left(2\mathrm{c_k^p} - 1\right) + \mathrm{A_{k-1}}\left(\mathrm{S_k^S(e)}\right) + \mathrm{B_k}(\mathrm{S_k^K(e)}) \right] \end{aligned} \tag{5}$$

$$\begin{aligned}\Lambda_e(u_k) =& \Lambda_k(u,o) - \Lambda_a(u_k) - \Lambda_k(c^s, I) \\ =& \max_{e,u(e=1)}\left[\frac{1}{2}\Lambda_k(c^s, I)\left(2c_k^p - 1\right) + A_{k-1}\left(S_k^S(e)\right) + B_k(S_k^K(e))\right] \\ &- \max_{e,u(e=0)}\left[\frac{1}{2}\Lambda_k(c^s, I)\left(2c_k^p - 1\right) + A_{k-1}\left(S_k^S(e)\right) + B_k\left(S_k^K(e)\right)\right]\end{aligned} \tag{6}$$

Since the use of FPGA in the decoding process for fixed-point operations, we must consider the possible overflow problem, in order to prevent the phenomenon of overflow, the need for various types of measurement process to normalize the process. The formula is normalized as follows:

$$\begin{aligned}A_k(s) =& \max_{e,S^S(E)}\left[A_{k-1}S_k^E(s) + M_k(E)\right] \\ &- \max_{S_e}\left[\max_{e,S^S(E)}\left[A_{k-1}S_k^E(s) + M_k(E)\right]\right], K = 1, 2, \ldots, N-1\end{aligned} \tag{7}$$

$$\begin{aligned}B_k(s) =& \max_{e,S^S(E)}\left[B_{k-1}S_{k+1}^E(s) + M_k(E)\right] - \max_{S_e}\left[\max_{e,S^S(E)}\left[A_k S_k^E(s) + M_{k+1}(E)\right]\right], \\ & K = 1, 2, \ldots, N-1\end{aligned} \tag{8}$$

The entire decoding process can be divided into four steps: initialization, data storage, iterative decoding and hard decision output. We use the state machine to control (Fig. 6).

The simulation results of Turbo Decoder are shown in the following Fig. 7.

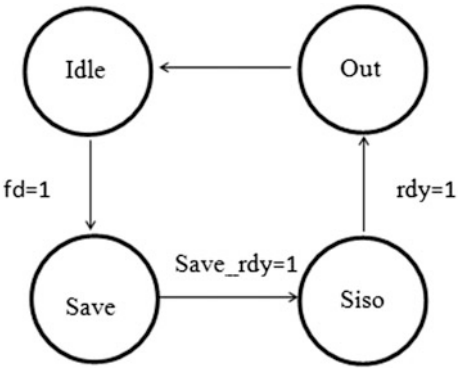


Fig. 6. Decoder state transition

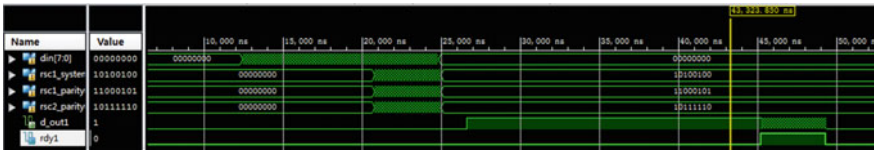


Fig. 7. Simulation results of Turbo decoder

6 Simulation Analysis and Comparison

We use Xilinx ISE tools to simulate the Turbo codes, decoding algorithm is Max-Log-MAP, interleaver length is 1024, S/N is 1.0 dB. The decoded data is compared with the original information data to get the wrong number of bits. The simulation results of 10 sets of data are shown in Table 1.

Table 1. Decoder error correction performance table

	Number of iterations				
	1 iterations	2 iterations	3 iterations	4 iterations	5 iterations
1.	75	58	25	2	2
2.	89	26	0	0	0
3.	62	31	10	4	1
4.	102	18	8	1	1
5.	96	60	34	1	0
6.	58	15	13	1	1
7.	94	66	25	18	9
8.	74	9	9	5	4
9.	83	22	20	15	8
10.	108	41	11	3	1
Total	841	346	159	49	27

Simulation of the same 10 sets of data using Matlab. Compare the bit error rate of ISE simulation and Matlab simulation.

Table 2. Compare the bit error rate of ISE simulation and Matlab simulation

	Number of iterations				
	1 iterations	2 iterations	3 iterations	4 iterations	5 iterations
ISE	0.0788	0.0312	0.0142	0.0054	0.0024
Matlab	0.0821	0.0338	0.0155	0.0055	0.0026

From Table 2 Turbo code in the FPGA performance can be obtained and Matlab simulation performance difference is very small. We can conclude that the FPGA design of this paper is correct and has good performance.

References

1. Goddemeier, N., Daniel, K., Wietfeld, C.: Coverage evaluation of wireless networks for unmanned aerial systems [C]. In: 2010 IEEE Globecom Workshops, vol. 3, pp. 1760–1765 (2010)
2. Auluzzi, D.R., Beaulieu, N.C.: A comparison of SNR estimation techniques for the AWGN channel [J]. IEEE Trans. Commun. **48**(10), 1681–1691 (2000)
3. Berrou, C., Glavieux, A., Thitimajshima, P.: Near shannon limit error-correcting codeing and decoding: turbo codes. In: Proceedings of ICC 1993, Geneva, Switzerland, pp. 1064–1070 (1993)
4. Robertson, P., Hoeher, P., Villebrun, E.: Optimal and sub optimal maximum aposteriori algorithm suitable for turbo decoding. Eur. Trans. Telecommun. **8**, 119–125 (1997)
5. Hagenauer, J., Robertson, P., Papke, L.: Iterative decoding of systematic convolutional codes with the MAP and SOVA algorithm. In: Proceedings of ITG 1994, pp. 1–9, October 1994
6. Chaikalis, C., Noras, J.M.: Implementation of an improved reconfigurable SOVA log- MAP turbo decoder in 3 GPP. In: Conference Publication on 3G Mobile Communication Technologies, vol. 489, 8–10 May 2002
7. Cheng, J.F., Ottosson, T.: Linearly approximated log-MAP algorithms for turbo decoding. In: Proceedings of IEEE Vehicular Technology Conference, Tokyo, Japan, 15–18 May 2000, pp. 2252–2256 (2000)

Channel Modelling of Dual-Polarized Mobile Satellite MIMO Communication System

Xian Xirui[1], Yang Mingchuan[2(✉)], Ma Chen[2], and Bi Ran[2]

[1] Measurement and Control Technology and Instrument, Harbin Institute of Technology, Harbin, China
[2] Communication Research Center, Harbin Institute of Technology, Harbin 150001, China
mcyang@hit.edu.cn

Abstract. In most terrestrial wireless network systems, MIMO (Multiple-Input Multiple-Output) technology can be used to make the system maintain a lower bit error rate and outage probability while gaining greater capacity gain, so more and more researchers have tried to apply MIMO technology to satellite mobile communication systems to alleviate the problem of increasingly crowded orbital locations and increasingly strained frequency resources. In this paper, the channel capacity, bit error rate and outage probability of 2 × 2 dual-polarized mobile MIMO communication systems are studied in-depth. First, when the antenna's XPD (Cross-Polarization Discrimination) is greater than 0, in a certain range, the channel capacity increases approximately linearly with the XPD increase. Second, the BER of the system under BPSK modulation is analyzed. When the SNR is large, the system can maintain a sufficiently low BER; an analysis of the outage probability of a distributed MIMO system shows the system can maintain a sufficiently low outage probability when the SNR is large. Taking into account the particularity of the satellite mobile communication channel, the influence factors of Rice channel are added in the analysis process. The analysis of these properties of the dual-polarized satellite mobile MIMO communication system will provide strong support for the future development of satellite MIMO technology.

Keywords: Dual-polarized satellite system · MIMO technology · Channel capacity · Outage probability · Cross-polarization discrimination

1 Introduction

With the continuous development of wireless communication technology, how to use a limited wireless resource to face a huge amount of communication data has become an important challenge in the current information field. Because MIMO technology can increase system spectrum utilization and channel capacity without increasing the system bandwidth and the total transmit power of the antenna, and thus fully utilize space resources, researchers have generally considered it as one of the most promising technologies for applying potential in wireless communication field in recent years. For example, deploying a large-scale array of large-scale multiple-input multiple-output (MIMO) technologies at the base station end cannot only effectively resist interference

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 274–284, 2019.
https://doi.org/10.1007/978-981-13-6264-4_33

between different users, but also significantly increase system capacity, and can also focus signal energy on very narrow beams. Effectively increase energy efficiency [1].

Although satellite mobile communications are in an indispensable position in the current global communications network, the increasingly serious problems such as the increasingly congested satellite orbits and the increasingly tense frequency resources make people eagerly hope that MIMO technology can also be successful in satellite mobile communications. The system capacity is increased under the condition of maintaining low bit error rate and outage probability to meet the growing user demand [2, 3]. The United States has already conducted in-depth studies in this area. The U.S. orbit diversity system has been successfully applied to digital audio wireless service satellites, Sirius, and XM satellite broadcasting in the United States. The European Telecommunications Standards Institute has also developed the DVB-SH standard for MIMO technology on mobile satellites [4].

The satellite mobile MIMO system is different from the terrestrial wireless MIMO system, which has inherent characteristics. When the MIMO technology is applied, design improvements are needed to fully consider the influence of large delay, small satellite payload, and unfixed position. At present, scholars mainly discuss the effects of satellite link atmospheric loss and rain attenuation, receiver and receiver settings, polarization method, and channel correlation on the capacity of satellite mobile MIMO channels. This paper takes the single satellite as the background, taking into account the particularity of the satellite mobile communication channel, and adding the influence factor of Rice channel, studies the channel capacity, bit error rate and outage probability of the dual-polarized satellite mobile MIMO system [5, 6].

2 Dual-Polarized Satellite Mobile MIMO System Model

The basic structure of a dual-polarized satellite mobile MIMO communication system is shown in Fig. 1, which constructs a 2 × 2 dual-polarized MIMO channel. Different polarized branch correlation coefficients are very low and fully meet the requirements

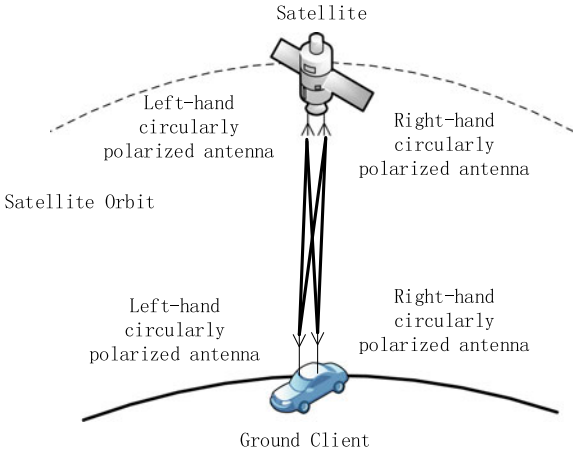


Fig. 1. Dual-polarized satellite mobile MIMO communication system

of diversity (ideally, the correlation coefficient is 0) [7]. Moreover, no matter how scatterers are distributed in dual-polarized channels, at least two independently transmitted parallel sub-channels can be obtained.

First, define the dual-polarized MIMO channel matrix as H. In this defined system, both the transmitter and the receiver must have even-numbered root antennas, and the average allocation is left-hand circular polarization and right-hand circular polarization. Considering that the cross-polarized components generated by other channels in the antenna system in practical applications will affect the normal communication of the channel operating at the same frequency, cross-polarization discrimination (XPD), that is, the main polarization component of the channel is introduced. The ratio of cross-polarized components generated by another channel within the channel. In the analytical model, define XPD as a scalar α, using matrix $\mathbf{A}$ to represent:

$$\mathbf{A} = \begin{pmatrix} 1 & \sqrt{\alpha} \\ \sqrt{\alpha} & 1 \end{pmatrix} \tag{1}$$

The Rice dual-polarized channel model can be expressed as follows:

$$\mathbf{H} = \mathbf{AH} = \sqrt{\frac{K}{1+K}}\mathbf{AH_{LOS}} + \sqrt{\frac{1}{1+K}}\mathbf{AH_{NLOS}} \tag{2}$$

A dual-polarized diversity MIMO channel matrix should contain the following information:

(a) The co-site antenna correlation should be 1 due to the correlation between antennas due to limited space;
(b) De-correlation between antennas due to polarization diversity;
(c) Power imbalance caused by co-polarized and inter-polarized transmissions.

3 Rice Channel Polarized MIMO Channel Capacity

When the scattering environment is abundant, the channels are mostly independent and identically distributed Rayleigh fading channels. However, if there is line-of-sight propagation between the transmitters and receivers, different fading statistics are generated, such as Rice fading channels. This is also more suitable for satellite MIMO. Channel in communication. In this paper, we study a MIMO system based on Rice fading channel, first consider a broadband indoor MIMO wireless transmission system composed of M_T transmitting antennas and M_R receiving antennas.

Assuming that the transmit channel has a flat frequency response, when the M_T dimensional transmit signal $\mathbf{x} = (x_1, x_2, \cdots, x_{M_T})^T$ reaches the receiving antenna by the channel, the received signal sequence $\mathbf{y} = (y_1, y_2, \cdots, y_{M_R})^T$, can be expressed by formula (3):

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon} \tag{3}$$

H is the channel matrix of $M_T \times M_R$ (in this article, only the case of $M_T = M_R$ is discussed), $\boldsymbol{\varepsilon}$ is a statistically independent zero mean, variance is 1 $M_R \times 1$ additive complex Gaussian white noise variable, and its correlation matrix $\mathbf{R}_{\varepsilon\varepsilon}$ satisfies the formula (4):

$$\mathbf{R}_{\varepsilon\varepsilon} = E\left\{\varepsilon\varepsilon^H\right\} = \sigma^2\mathbf{I}_{M_R} \tag{4}$$

Assume that the element h_{ji} of the channel matrix satisfies:

$$h_{ji} = r_{ji}e^{j\varphi}\mathrm{Re}\left(h_{ji}\right) + j\mathrm{Im}\left(h_{ji}\right) \tag{5}$$

Both $\mathrm{Re}\left(h_{ji}\right)$ and $\mathrm{Im}\left(h_{ji}\right)$ are Gaussian stochastic processes with a variance of Ω, and the mean values are not zero, Re(A) and Im(A), respectively. The mean of h_{ji} is not zero, the envelope follows the Rice distribution, and the phase obeys the uniform distribution of [0, 2π).

In Rice fading, the ratio K of the power of the line-of-sight component and the scattered component is generally used as the main parameter of the Rice distribution, known as Rice, which is usually expressed in dB:

$$K = \frac{A^2}{2\Omega} \tag{6}$$

The channel is typically modelled as the sum of the LOS component and the non-LOS component (NLOS). As shown in Eq. (7).

$$\mathbf{H}(f) = \sqrt{\frac{K}{1+K}}\mathbf{H}_{\mathrm{LOS}}(f) + \sqrt{\frac{1}{1+K}}\mathbf{H}_{\mathrm{NLOS}}(f) \tag{7}$$

The first term on the right side of the equation is the line-of-sight part of the channel, and the second term is the channel on the assumption that the channel has no relevant fading. $\mathrm{H_{LOS}}$ is a completely deterministic matrix, and $\mathrm{H_{NLOS}}$ is a random flat fading component. It is a channel matrix which assumes that the system has only scatter fading conditions. It is a random zero-mean complex Gaussian matrix.

Take the transmitter unknown channel state information as an example. At this time, the MIMO capacity satisfies Eq. (8):

$$C = W\mathrm{m}log_2\left(1 + \frac{P}{\sigma^2}\right) \tag{8}$$

Among them, C is the system channel capacity, W is the channel bandwidth, P/σ^2 is the signal-to-noise ratio, $\mathrm{m} = \min(M_R, M_T)$.

The following describes several important parameters used to calculate the capacity of a MIMO system with Rice fading.

3.1 Line-Of-Sight (LOS) Matrix

From M_T transmit antennas, get M_T $\mathrm{H_{LOS}}$ column vectors:

$$\mathrm{H_{LOS}} = \left[h^{-(1)}, h^{-(2)}, \cdots, h^{-(M_T)}\right] \tag{9}$$

Take $M = M_T = M_R$, there is the following formula:

$$h^{-(m)} = \left[1, e^{-j\beta_m}, \cdots, e^{-j(M-1)\beta_m}\right]^T \tag{10}$$

Among them, β_m is the phase change of the incident wave to the receiving antenna. m traverses from 1 to M.

If two dual-polarized antennas are used at the transmitting end and the receiving end instead of four single-polarized antennas, the system becomes a 2×2 dual-polarized MIMO system, and the matrix $\mathrm{H_{LOS}}$ is obtained by the following equation:

$$\mathrm{H_{LOS}} = \left[h^{-(1)}, h^{-(2)}, \cdots, h^{-(M_T)}\right] \otimes \mathrm{I}_{(2,2)} \tag{11}$$

Among them, $M = M_T = M_R$, $I_{(2,2)}$ is a 2-order all-one matrix.

3.2 Polarized Correlation Matrix

The correlation matrix of a MIMO system is often divided into two parts: spatial correlation matrix and polarized correlation matrix. In order to analyze the influence of polarization characteristics, the polarization correlation matrix is considered in this paper. The use of antennas with different polarizations at the transmitter and receiver ends may result in power and related imbalances between MIMO channels. The degree of correlation between polarized antennas can be quantified using XPR (Cross-polarization ratio). The XPR is the power ratio between the cross-polarized antenna pair and the co-polarized antenna pair.

The antenna polarization matrix defining the transmitter and receiver is:

$$\mathrm{S} = \begin{pmatrix} S_{ll} & S_{lr} \\ S_{rl} & S_{rr} \end{pmatrix} \tag{12}$$

In the formula, the subscript l indicates left circular polarization, and r indicates right circular polarization. The first subscript indicates the polarization of the transmitter and the second subscript indicates the polarization of the receiver. Then there are

$$\mathrm{XPR} = \frac{S_{lr}}{S_{ll}} = \frac{S_{rl}}{S_{rr}} \tag{13}$$

According to [8], the cross-polarization ratio depends only on the polarization state of both the antenna under test and the test antenna, and the XPR depends on the polarization scheme and the propagation path of the transmitting and receiving

antennas. In the ideal case, the gain represented by XPR is zero. Considering non-ideal conditions, XPD (dB) needs to be considered for cross-polarization, with a range of (−30, 30), and defines α as

$$\mathrm{XPD} = 10\log_{10}[(1-\alpha)/\alpha] \tag{14}$$

The dual-polarized channel model can be expressed as follows:

$$\mathbf{H} = \mathbf{AH} = \begin{pmatrix} 1 & \sqrt{\alpha} \\ \sqrt{\alpha} & 1 \end{pmatrix} \otimes \mathbf{H} \tag{15}$$

3.3 Non-Line-Of-Sight (NLOS) Matrix

The H_{NLOS} component in Eq. (7) is a random flat fading part of the system. Suppose an $M_T \times M_R$ matrix G represents an uncorrelated flat fading channel coefficient, then a random component H_{NLOS} can be represented by

$$vec\{\mathbf{H}_{\mathrm{NLOS}}\} = \sqrt{\mathbf{R}} \times vec\{\mathbf{G}\} \tag{16}$$

where R is the total correlation matrix of the transmitting antenna and the receiving antenna.

3.4 Antenna Selection

The omnidirectional antenna appears to be uniformly irradiating in the 360° direction in the horizontal direction, that is, it has no directionality. Directional antennas are particularly strong in transmitting and receiving electromagnetic waves in one or a few specific directions, and are zero or extremely small in transmitting and receiving electromagnetic waves in other directions. The purpose of using a directional transmit antenna is to increase the effective utilization of radiation power, confidentiality, and anti-jamming capability.

Combined with the above principles, we can consider that when the MIMO system uses a directional antenna, its line-of-sight signal will be strengthened, and the NLOS signal will be weakened, which has the same effect as the Rice factor *K*.

4 BPSK Demodulation Polarized MIMO Bit Error Rate

The bit error rate is also an important performance measure of a digital communication system. Under the interference of Gaussian white noise in the channel, the error rate of various binary digital modulation systems depends on the demodulator input SNR, and the form of the BER expression depends on the demodulation method. In this paper, common BPSK (Binary Phase Shift Keying) demodulation method is used to analyze the BER of 2 × 2 dual-polarized satellite MIMO communication system.

Make the signal sent by the sending end be as shown in the formula, then the output band-pass filter output waveform y(t) is

$$y(t) = \begin{cases} [a + n_c(t)]\cos\omega_c t - n_s(t)\sin\omega_c t \; sending ``1" \\ [-a + n_c(t)]\cos\omega_c t - n_s(t)\sin\omega_c t \; sending ``0" \end{cases} \tag{17}$$

After y(t) has been coherently demodulated (multiply–low pass), the input waveform sent to the sampler is

$$x(t) = \begin{cases} a + n_c(t) & send\ symbol ``1" \\ -a + n_c(t) & send\ symbol ``0" \end{cases} \tag{18}$$

Since $n_c(t)$ is a Gaussian noise with a mean of 0 and a variance of σ_n^2, the one-dimensional probability density function of x(t) is

$$f_1(x) = \frac{1}{\sqrt{2\pi}\sigma_n} exp\left\{ -\frac{(x-a)^2}{2\sigma_n^2} \right\} sending ``1" \tag{19}$$

$$f_0(x) = \frac{1}{\sqrt{2\pi}\sigma_n} exp\left\{ -\frac{(x+a)^2}{2\sigma_n^2} \right\} sending ``0" \tag{20}$$

According to the analysis of the optimal decision threshold, when the probability of sending a "1" symbol and sending a "0" symbol is equal, that is, P(1) = P(0), the optimal decision threshold $b^* = 0$. At this time, the probability of "1" is issued, and the erroneous judgment is "0".

$$P(0/1) = P(x \le 0) = \int_{-\infty}^{0} f_1(x)dx = \frac{1}{2} erfc\left(\sqrt{r}\right) \tag{21}$$

In the formula $r = \frac{a^2}{2\sigma_n^2}$

Similarly, the probability of sending "0" and erroneously determining "1" is

$$P(1/0) = P(x \le 0) = \int_{0}^{\infty} f_0(x)dx = \frac{1}{2} erfc\left(\sqrt{r}\right) \tag{22}$$

Therefore, the total bit error rate during coherent demodulation of BPSK signals is

$$P_e = P(1)P(0/1) + P(0)P(1/0) = \frac{1}{2} erfc\left(\sqrt{r}\right) \tag{23}$$

5 Polarized MIMO Outage Probability

The bipolar MIMO system outage probability is defined as the probability that the system output SNR is below a certain value. In this paper, given the equivalence of the received SNR, the outage probability at this time can be expressed as

$$P_{out} = F_{\gamma}(\gamma_{th}) = \prod_{i=1}^{K} \left[1 - \frac{1}{2} erfc\left(\frac{10lg\gamma_{th} - \hat{\mu}_i}{\sqrt{2}\hat{\sigma}_i} \right) \right] \tag{24}$$

γ_{th} is the SNR threshold.

6 Result Simulation and Analysis

In the simulation, the system model shown in Fig. 1 is used, where $M_R = M_T = 2$, the receiver elevation angle $\theta_R = 13°$, and the azimuth angle $\phi_R = 52°$. However, due to the fact that the XPD of the satellite-side antenna is usually very large, only the antenna *XPD* of the terrestrial receiver is considered here. Figure 2 is the simulation result of the capacity of the dual-polarized satellite mobile MIMO system changing with the ground receiver *XPD*.

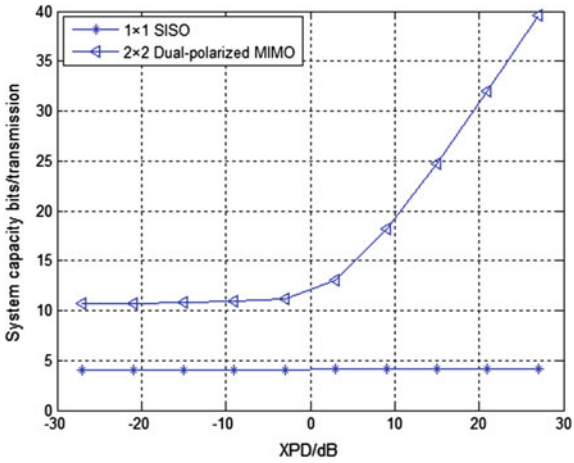


Fig. 2. Influence of XPD on the capacity of dual-polarized satellite MIMO system

It can be seen that the situation is different before and after XPD = 0 dB. Before 0, the channel capacity is hardly affected by the XPD. At this time, the ratio of the main polarization component of the channel to the cross-polarization component generated in the other channel in the channel is close to 0, and the main polarization component becomes negligible. After 0, the cross-polarization component produced by the other channel in this channel is getting smaller and smaller, the diversity effect is getting better and better, and the channel capacity increases approximately linearly with the increasement of *XPD*.

Figure 3 shows the capacity comparison of a distributed and dual-polarized MIMO system. $M_R = M_T = 2$, the receiver elevation angle $\theta_R = 13°$, azimuth angle $\phi_R = 52°$, *XPD* = 30 dB.

It can be seen from the figure that the distributed satellite mobile MIMO system has a larger capacity gain than the dual-polarized satellite mobile MIMO system, but the dual-polarized MIMO system can usually overcome the problems of delay and synchronization. In short, the satellite mobile MIMO system formed by any method is always higher than the channel capacity of the SISO system.

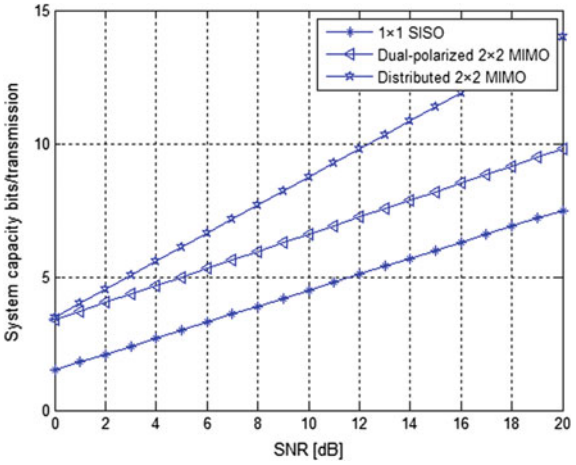


Fig. 3. Comparison of capacity between distributed and dual-polarized MIMO systems

Figure 4 shows the BER variation of the dual-polarized MIMO system in BPSK demodulation mode. It can be seen from the figure that the derived signal-to-noise ratio curve of BER almost completely coincides with the curve obtained from simulation, showing that the theoretical analysis should be correct. With the increase of SNR, the bit error rate of the dual-polarized satellite mobile MIMO system can fully reach 10^{-6}, which can be applied to practical projects.

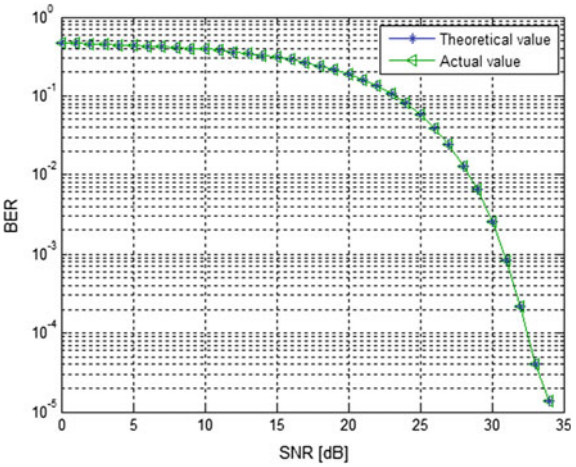


Fig. 4. BPSK demodulated polarization MIMO bit error rate

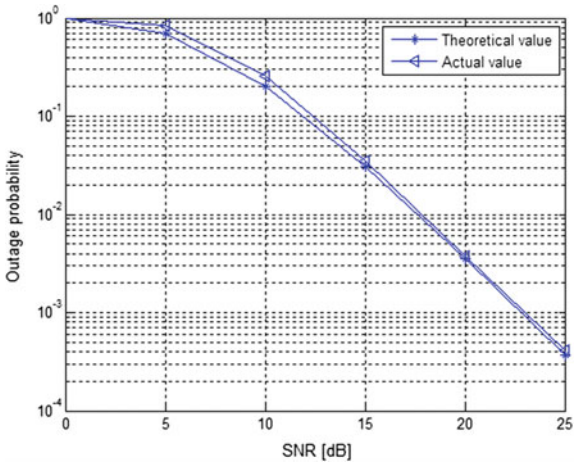


Fig. 5. Dual-polarized MIMO outage probability

Figure 5 shows the relationship between the outage probability and signal-to-noise ratio of a dual-polarized MIMO communication system. The simulation curve and the theoretical curve are basically the same, and the system interruption probability shows a declining trend as the signal-to-noise ratio gradually increases. When the SNR is large enough, the outage probability of a dual-polarized MIMO communication system can fully meet the actual engineering requirements.

7 Conclusion

In this paper, MIMO technology is introduced into the satellite mobile communication system by means of polarization diversity, and the polarized MIMO capacity, bit error rate and outage probability of Rice fading channel are calculated. It can be seen that in a dual-polarized satellite mobile MIMO system, polarization diversity can significantly increase the system capacity compared to a single-input single-output system while ensuring that the system bit error rate and outage probability are sufficiently low, and its improvement range increases with XPD increasing.

Acknowledgment. The paper is sponsored by National Natural Science Foundation of China (No. 91538104; No. 91438205).

References

1. Arapoglou, P.-D., Liolis, K., Bertinelli, M., Panagopoulos, A., Cottis, P., De Gaudenzi, R.: MIMO over satellite: a review. IEEE Commun. Surv. Tutor. **13**(1), 27–51 (2011)
2. Paulraj, A.J., Gore, D.A., Nabar, R.U., Bolcskei, H.: An overview of MIMO communications - a key to gigabit wireless. Proc. IEEE **92**(2), 198–218 (2004)

3. Arapoglou, P.-D., Burzigotti, P., Bertinelli, M., Bolea-Alamanac, A., De Gaudenzi, R.: To MIMO or not to MIMO in mobile satellite broadcasting systems. IEEE Trans. Wirel. Commun. **10**(9), 2807–2811 (2011)
4. Kelley, P.: Overview of the DVB-SH specifications. Int. J. Satell. Commun. Netw. **27**(4–5), 193–293 (2009)
5. Kyrolainen, J., Hulkkonen, A., Ylitalo, J., Byman, A., Shankar, B., Arapoglou, P., Grotz, J.: Applicability of MIMO to Satellite Communications, pp. 343–357. Wiley (2013)
6. Perez-Neira, A.I., Ibars, C., Serra, J., del Coso, A., Gomez-Vilardebo, J., Caus, M., Liolis, K. P.: MIMO channel modeling and transmission techniques for multi-satellite and hybrid satellite-terrestrial mobile networks. Phys. Commun. **4**, 127–139 (2011)
7. Jensen, M.A., Wallace, J.W.: A review of antenna sand propagation for MIMO wireless communications. IEEE Trans. Antennas Propag. **52**(11) (2004)
8. Xianrong, S., Yuliang, D., Hua, W.: The cross-polarization ratio of the antenna. J. Chin. Acad. Sci. **5**(4), 344–346 (2008)

Rate Analysis of Quantized MIMO Systems with Transmit Channel State Information

Xiaofeng Ling[1(✉)], Rui Wang[2], and Zhaozhao Gao[3]

[1] The School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, People's Republic of China
xfling@ecust.edu.cn

[2] The Department of Information and Communications, Tongji University, Shanghai 201804, People's Republic of China

[3] Science and Technology on Electronic Information Control Laboratory, Chengdu, People's Republic of China

Abstract. The achievable rate of the multi-input multi-output (MIMO) channel with low-resolution receiver ADCs is analyzed. With the channel state information at both the transmitter and receiver, we jointly optimize the transmit signal and the receiving analog combiner with an aim to improve the performance. To this goal, the approximate achievable rate of considered channel model is first obtained using the Bussgang theorem. After that, the singular-value decomposition (SVD) based approach is proposed to enable the joint optimization. Our simulation results show that the proposed design is able to reach the capacity at low signal-to-noise ratio (SNR).

Keywords: Multi-input Multi-output (MIMO) · Quantization · Achievable Rate

1 Introduction

Due to the large channel bandwidth, the millimeter wave (mmWave) carrier frequency band has a great potential to improve the transmission rate of cellular networks. Comparing with the transmit signals over low frequency bands, the transmit signals suffer larger attenuations in mmWave frequency bands. In this case, using antenna array can be considered as a promising way to improve the performance of mmWave systems.

Transmission in higher carrier frequencies brings some design problems. One major issue is that the high sampling rate with high-resolution analog-to-digital conversions (ADCs) brings large power consumption. Moreover, the power cost increases more significantly when with more antennas. As it has been proven that the power consumption grows exponentially with the resolution of ADCs [1], one promising solution to reduce the power consumption is to user low-resolution ADCs.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 285–291, 2019.
https://doi.org/10.1007/978-981-13-6264-4_34

Low-resolution ADCs modifies the established traditional communication theories and designs. The corresponding studies have attracted wide attention. The single-input single-output (SISO) channel with one-bit ADCs was studied in [2]. The authors in [2] showed that when knowing transmit channel state information (CSIT), binary antipodal signaling was optimal. The authors in [3] proved that when without CSIT, the QPSK signaling is optimal for block fading SISO channel. The multi-input multi-output (MIMO) technique was considered in [4–6]. The authors in [4] studied the capacity of one-bit quantized MIMO channels where the authors derived the exact capacity of the multiple-input single-output channel. In [5], the authors considered the massive MIMO multi-user uplink channel. When without the CSI at both the transmitter and receiver, the linear minimum mean square error (LMMSE) based channel estimation was investigated and the achievable rate was analyzed. In [6], the authors analyzed the achievable rate with hybrid beamforming.

In this work, we focus on optimizing the achievable rate of the MIMO channel model with low-resolution ADC with few-bit quantization. With the channel state information at both the transmitter and receiver, the transmit signal and the receiving analog combiner are jointly optimized with an aim to enhance the achievable rate. To achieve this goal, we derive the approximate achievable rate by utilizing the Bussgang theory. To improve the achievable rate, a singular-value decomposition (SVD) based approach is proposed to perform the joint optimization. Our simulation results show that the proposed design is able to reach the capacity at low signal-to-noise ratio (SNR).

Notations: $\mathbb{E}(\cdot)$ denotes the expectation operator. Superscripts $\mathbf{A}^T$, $\mathbf{A}^*$, and $\mathbf{A}^H$ denote the transpose, conjugate, and conjugate transpose of matrix $\mathbf{A}$, respectively. $\mathrm{Tr}(\mathbf{A})$, $\mathbf{A}^{-1}$ and $\det(\mathbf{A})$ denote the trace, inverse and determinant of $\mathbf{A}$, respectively. $\mathrm{diag}(\mathbf{a})$ denotes a diagonal matrix with $\mathbf{a}$ being its diagonal entries. $\mathbf{0}$ and $\mathbf{I}$ denote the zero and identity matrices, respectively. The distribution of a circular symmetric complex Gaussian vector with mean vector $\mathbf{x}$ and covariance matrix $\mathbf{\Sigma}$ is denoted by $\mathcal{CN}(\mathbf{x}, \mathbf{\Sigma})$. $\mathbf{A} \succeq \mathbf{0}$ implies that matrix $\mathbf{A}$ is a semidefinite positive matrix.

2 System Model

In the considered MIMO channel, we assume that there are n and m antennas at the transmitter and the receiver, respectively. The received signal in the system is given by

$$\mathbf{y}' = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{H}$ represents the channel matrix, $\mathbf{x}$ denotes the transmit signal, $\mathbf{y}'$ denotes the received signal before post-processing, and $\mathbf{n}$ denotes the additive noise following $\mathcal{CN}(\mathbf{0}, \mathbf{I})$. Additionally, signal $\mathbf{x}$ satisfies the power constraint of

$$\mathrm{tr}(\mathbb{E}(\mathbf{x}\mathbf{x}^H)) = \mathrm{tr}(\mathbf{Q}) \leq P_t \tag{2}$$

where $\mathbf{Q} = \mathbb{E}(\mathbf{x}\mathbf{x}^H)$ is the signal covariance matrix and P_t denotes the power target. The singular-value decomposition of the channel $\mathbf{H}$ is represented by

$\mathbf{H} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^H$ where $\mathbf{U}$ and $\mathbf{V}$ are $m \times m$ and $n \times n$ unitary matrices, respectively, $\boldsymbol{\Lambda}$ is a diagonal matrix with its diagonal elements $\{\lambda_1, \lambda_2, \cdots, \lambda_{\min(n,m)}\}$ being arranged in decreasing order and λ_i being the i-th singular value.

Before inputting the signal to ADC, the received signal $\mathbf{y}'$ is multiplied by a analog combiner $\mathbf{W}$. We have

$$\mathbf{y} = \mathbf{W}\mathbf{y}' = \mathbf{W}\mathbf{H}\mathbf{x} + \tilde{\mathbf{n}}. \tag{3}$$

where $\tilde{\mathbf{n}} = \mathbf{W}\mathbf{n}$. Then, signal $\mathbf{y}$ is quantized by $2mb$ bits with each streams being separately quantized by $2b$ bits where b bits are for real part and the remaining b bits are for imaginary part. The output signal is denoted by

$$\begin{aligned} \mathbf{r} &= Q(\mathbf{y}) \\ &= Q(\mathbf{W}\mathbf{y}') \\ &= Q(\mathbf{W}\mathbf{H}\mathbf{x} + \mathbf{W}\mathbf{n}). \end{aligned} \tag{4}$$

Using the Bussgang theorem to linearly model the quantized signal $\mathbf{r}$, the achievable rate can be represented as

$$\mathbf{r} = \mathbf{F}\mathbf{y} + \mathbf{e} \tag{5}$$

where $\mathbf{e}$ is the uncorrelated distortion, and $\mathbf{F}$ can be obtained from the linear MMSE estimation of $\mathbf{r}$ from $\mathbf{y}$, given by

$$\mathbf{F} = \mathbb{E}(\mathbf{r}\mathbf{y}^H)\mathbb{E}(\mathbf{y}\mathbf{y}^H)^{-1} = \mathbf{R}_{ry}\mathbf{R}_{yy}^{-1}. \tag{6}$$

By considering $\mathbf{WH}$ as equivalent channel and $\mathbf{Wn}$ as equivalent additive noise, the achievable rate can be approximated as

$$\begin{aligned} r = &\log|\mathbf{I} + (1-\rho_q)\left((1-\rho_q)\mathbf{R}_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}} + \rho_q \mathrm{diag}(\mathbf{R}_{\mathbf{yy}})\right)^{-1} \\ &\times \mathbf{WHQH}^H\mathbf{W}^H| \end{aligned} \tag{7}$$

where ρ_q is the quantization distortion error determined by the quantizer and the number of bits, $\mathbf{R}_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}} = \mathbb{E}\left(\tilde{\mathbf{n}}\tilde{\mathbf{n}}^H\right) = \sigma^2\mathbf{W}\mathbf{W}^H$, $\mathbf{R}_{\mathbf{yy}} = \sigma^2\mathbf{W}\mathbf{W}^H + \mathbf{WHQH}^H\mathbf{W}^H$.

3 Joint Optimization of Transmit Signal and Receive Analog Combiner

The transmit signal optimization is actually to optimize the statistics of the signal $\mathbf{x}$, i.e., $\mathbf{Q}$. Hence, the joint design problem can be represented as

$$\begin{aligned} \max_{\mathbf{Q},\mathbf{W}} \quad & \log|\mathbf{I} + (1-\rho_q)\left((1-\rho_q)\sigma^2\mathbf{W}\mathbf{W}^H + \right. \\ & \left.\rho_q \mathrm{diag}(\sigma^2\mathbf{W}\mathbf{W}^H + \mathbf{WHQH}^H\mathbf{W}^H)\right)^{-1} \\ & \times \mathbf{WHQH}^H\mathbf{W}^H| \qquad (8a) \\ \text{s.t.} \quad & \mathrm{Tr}(\mathbf{Q}) \le P_t \qquad (8b) \end{aligned}$$

In what follows, we consider to use a special structures of $\mathbf{Q}$ and $\mathbf{W}$ to simplify the design. To parallelize the MIMO channel, we design $\mathbf{Q}$ and $\mathbf{W}$ such that $\mathbf{W}\mathbf{W}^H$ and $\mathbf{W}\mathbf{H}\mathbf{Q}\mathbf{H}^H\mathbf{W}^H$ in (8a) are diagonalized. According to the SVD decomposition of $\mathbf{H}$, we propose to user the following solution

$$\mathbf{Q} = \mathbf{V}\boldsymbol{\Lambda}_Q\mathbf{V}^H \text{ and } \mathbf{W} = \mathbf{U}^H \tag{9}$$

where $\mathbf{V}$ and $\mathbf{U}$ are unitary matrices related to the SVD decompostion of the channel, $\boldsymbol{\Lambda}_Q = \text{Diag}\,(\lambda_{q,1}, \lambda_{q,2}, \ldots, \lambda_{q,n})$. With $\mathbf{Q}$ and $\mathbf{W}$ given in (9), the channel is parallelized and the achievable rate in (6) depends on the power allocated to each sub-channel. In (9), the diagonal matrix $\boldsymbol{\Lambda}_Q$ and $\boldsymbol{\Lambda}_W$ are used to balance the power at the sub-channels.

Then, optimization problem in (8) reduces to

$$\min_{\lambda_{q,i}\geq 0} \; -\sum_{i=1}^{n} \log\left(1 + \frac{(1-\rho_q)\lambda_i^2\lambda_{q,i}}{\sigma^2 + \rho_q\lambda_i^2\lambda_{q,i}}\right) \tag{10a}$$

$$\text{s.t.} \; \sum_{i=1}^{n} \lambda_{q,i} \leq P_t \tag{10b}$$

Lemma 1. *Optimization problem* (10) *is convex and the optimal solution of* $\boldsymbol{\Lambda}_Q$ *is given by*

$$\begin{aligned}
\lambda_{q,i} &= \max\left[0, \frac{-\beta\sigma^2(\rho_q\lambda_i^2 + \lambda_i^2)}{2\beta\rho_q\lambda_i^4}\right. \\
&\left. + \frac{\sqrt{\beta^2\sigma^4(\rho_q\lambda_i^2 + \lambda_i^2)^2 - 4\beta\rho_q\lambda_i^4\left(\beta\sigma^4 - \frac{1}{\ln 2}\sigma^2(1-\rho_q)\lambda_i^2\right)}}{2\beta\rho_q\lambda_i^4}\right]^+ \\
&= \max\left[0, \frac{-\sigma^2(\rho_q + 1)}{2\rho_q\lambda_i^2}\right. \\
&\left. + \frac{\sqrt{\sigma^4(1-\rho_q)^2 + \frac{4}{\beta\ln 2}\sigma^2\rho_q(1-\rho_q)\lambda_i^2}}{2\rho_q\lambda_i^2}\right]^+
\end{aligned} \tag{11}$$

where β *is chosen to meet* $\sum_{i=1}^{n}\lambda_{q,i} = P_t$.

Proof. As the power constraint (10b) is linear, to prove the convexity of problem (10), we only need to prove that the objective function is concave. By representing $r = f(g(\lambda_i)$ with $f(x) = \log(x)$ and $g(\lambda_i) = 1 + \frac{(1-\rho_q)\lambda_i^2\lambda_{q,i}}{(1-\rho_q)\sigma^2 + \rho_q\sigma^2 + \rho_q\lambda_i^2\lambda_{q,i}}$, since $f(x)$ is a nondecreasing concave function and $g(\lambda_i)$ is a concave function with λ_i, the compositional function $f(\lambda_i)$ is concave [7].

Then, the optimal solution can be found by using KKT conditions where the Lagrangian function of (10) is given by

$$\begin{aligned}\mathcal{L} = &-\sum_{i=1}^{n}\log\left(1+\frac{(1-\rho_q)\lambda_i^2\lambda_{q,i}}{\sigma^2+\rho_q\lambda_i^2\lambda_{q,i}}\right)\\ &+\beta(\sum_{i=1}^{n}\lambda_{q,i}-P_t)-\sum_{i=1}^{n}\beta_i\lambda_{q,i},\end{aligned} \tag{12}$$

where β and β_i are the Lagrangian multipliers. With (12), the KKT conditions are given as

$$\frac{\partial\mathcal{L}}{\partial\lambda_{q,i}} = -\frac{1}{\ln 2}\frac{(1-\rho_q)\sigma^2\lambda_i^2}{(\sigma^2+\lambda_i^2\lambda_{q,i})(\sigma^2+\rho_q\lambda_i^2\lambda_{q,i})}+\beta-\beta_i = 0 \tag{13a}$$

$$\beta(\sum_{i=1}^{n}\lambda_{q,i}-P_t) = 0 \tag{13b}$$

$$\beta_i\lambda_{q,i} = 0 \tag{13c}$$

$$\beta \geq 0 \tag{13d}$$

$$\beta_i \geq 0, \forall i. \tag{13e}$$

Multiplying (13b) by $\lambda_{q,i}$, we have

$$\lambda_{q,i}\left(\beta-\frac{1}{\ln 2}\frac{(1-\rho_q)\sigma^2\lambda_i^2}{(\sigma^2+\lambda_i^2\lambda_{q,i})(\sigma^2+\rho_q\lambda_i^2\lambda_{q,i})}\right) = \lambda_{q,i}\beta_i = 0, \tag{14}$$

where the second equation is obtained with (13c). To satisfying (14), we consider the following two cases:

Case 1: If $\beta \geq \frac{1}{\ln 2}\frac{(1-\rho_q)\lambda_i^2}{\sigma^2}$, we must have $\lambda_{q,i} = 0$.

Case 2: Else, if $\beta < \frac{1}{\ln 2}\frac{(1-\rho_q)\lambda_i^2}{\sigma^2}$, as $\beta_i \geq 0$ in (13a), to satisfy (13a), we must have $\lambda_{q,i} > 0$. Furthermore, to meet (14), we must have $\beta = \frac{1}{\ln 2}\frac{(1-\rho_q)\sigma^2\lambda_i^2}{(\sigma^2+\lambda_i^2\lambda_{q,i})(\sigma^2+\rho_q\lambda_i^2\lambda_{q,i})}$.

In Case 2, the optimal $\lambda_{q,i}$ is obtained by solving

$$\beta\rho_q\lambda_i^4\lambda_{q,i}^2+\theta_1\lambda_{q,i}+\theta_2 = 0, \tag{15}$$

where $\theta_1 = \beta\sigma^2(\rho_q\lambda_i^2+\lambda_i^2)$ and $\theta_2 = \beta\sigma^4-\frac{1}{\ln 2}\sigma^2(1-\rho_q)\lambda_i^2$. The solution of (15) is given by

$$\lambda_{q,i} = \frac{-\theta_1+\sqrt{\theta_1^2-4\theta_2\beta\rho_q\lambda_i^4}}{2\beta\rho_q\lambda_i^4}. \tag{16}$$

Combining Cases 1 and 2, we obtain the final solution given in (11). From (13a), we observe that β must be larger than zero. Hence, to meet (13b), the Lagrangian multiplier β in (11) must satisfy $\sum_{i=1}^{n}\lambda_{q,i} = P_t$. We complete the proof of Lemma 1.

4 Simulation Results

Now we present simulation results to verify the proposed designs. The channels are modeled as Rayleigh fading where the elements of the channel matrix are complex Gaussian random variables with zero mean and unit variance. Since the noise powers are set to one, the SNR of the system can be represented as $\mathrm{SNR} = P_t$.

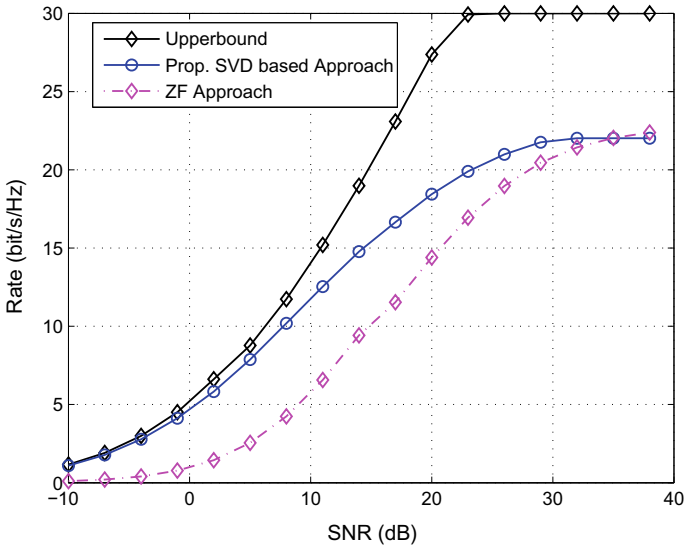


Fig. 1. Performance comparison of different designs with $n = m = 5$ and $b = 3$.

In Fig. 1, we compare the performance of different designs at $n = m = 5$ with $b = 3$. In zero-forcing (ZF) approach, the ZF beamformer proposed in [4] is used and assume $\mathbf{Q} = \mathbf{F}_{ZF}\mathbf{F}_{ZF}^H$ where $\mathbf{F}_{ZF} = \beta\mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}$ and β is scalar used to satisfy the power constraint. The curves in (1) implies that the proposed SVD based approach greatly outperforms the ZF approach. By comparing with the upper bound, we observe that at low SNR, the proposed SVD based approach are able to achieve the upper bound. With the increase of b, the SNR range approaching the upper bound becomes large.

5 Conclusion

By jointly optimizing the transmit signal and the receiving analog combiner, we studied the achievable rate optimization of the MIMO channel model with low-resolution ADCs at the receiver. The singular-value decomposition based approach was proposed to perform jointly optimization. By comparing with the upper bound, we showed that the upper bound can be reached at low SNR, which implies that the capacity can be achieved.

Acknowledgement. This work was supported by Science and Technology on Electronic Information Control Laboratory and by the National Science Foundation China under Grant 61771345.

References

1. Walden, R.H.: Analog-to-digital converter survey and analysis. IEEE J. Sel. Areas Commun. **17**(4), 539–550 (1999)
2. Singh, J., Dabeer, O., Madhow, U.: On the limits of communication with low-precision analog-to-digital conversion at the receiver. IEEE Trans. Commun. **57**(12), 3629–3639 (2009)
3. Mezghani, A., Nossek, J.A.: Analysis of Rayleigh-fading channels with 1-bit quantized output. In: 2008 IEEE International Symposium on Information Theory, pp. 260–264. Toronto, ON (2008)
4. Mo, J., Heath, R.W.: Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information. IEEE Trans. Signal Process. **63**(20), 5498-5512 (2015)
5. Jacobsson, S., Durisi, G., Coldrey, M., Gustavsson, U., Studer, C.: Throughput analysis of massive MIMO uplink with low-resolution ADCs. IEEE Trans. Wirel. Commun. **16**(6), 4038–4051 (2017)
6. Roth, K., Nossek, J.A.: Achievable rate and energy efficiency of hybrid and digital beamforming receivers with low resolution ADC. IEEE J. Sel. Areas Commun. **35**(9), 2056–2068 (2017)
7. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)

Improved Bit Allocation Using Distortion for the CTU-Level Rate Control in HEVC

Xin Lu[1], Bixing Zhou[1(✉)], Xuesong Jin[2], and Yanfeng Gu[1]

[1] School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China
374905099@qq.com
[2] School of Computer & Information Engineering, Harbin University of Commerce, Harbin 150028, China

Abstract. The research on rate control has always been a hot topic in video coding. Although H.265/HEVC is the latest video coding standard, the mismatch between the number of target bits and that of actual coding bits in the coding process still exists. A large number of experimental results have shown that the R-D model is superior to the R-λ model in terms of the matching degree of bit allocation. In this paper, a novel bit allocation method is proposed. First, the recursive Taylor expansion equation solving method is used to solve the rate-distortion (constrained problem) optimization equation, and then the R-D model is used to optimize the CTU-level target bit allocation.The simulation results indicate that the average bitrate error is only 0.19%, while the mean peak signal-to-noise ratio (MPSNR) increases by 0.14 dB.

Keywords: HEVC · R-D model · Rate control · Bit allocation

1 Introduction

Rate control is one of the core technologies in video coding, especially in the applications of real-time communication, as video coding has very strict requirements on channel bandwidth. Rate control is mainly divided into two parts: one is bit allocation, which allocates the appropriate number of bits for each coding level under the constraint of limited bandwidth; the other is the calculation of the Quantization Parameter (QP). The appropriate QP can reduce the coding error of rate control. In proposal M0036 [1] that was submitted to the Joint Collaborative Team on Video Coding (JCT-VC) [2], Li et al. used the R-λ model to realize a basic idea of bit allocation optimization. However, this method has not considered the problem that the number of bits actually encoded in the CTU (Coding Tree Unit) level does not match with that of the target bits. Therefore, the target bit allocation is a key for the study of rate control.

In rate control, the bit budget can be allocated to the Group of Pictures (GOP) level, frame level or CTU level, and it depends on the level at which the rate control is applied [3]. The weight of target bit allocation at the CTU level is based on the complexity of the coding unit. In most cases, the complexity was represented by the Mean Absolute Difference (MAD). In fact, the calculation of the quantization parameter is also important. In the proposal K0103 [4], Li et al. proposed a R-λ model

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 292–301, 2019.
https://doi.org/10.1007/978-981-13-6264-4_35

to calculate the quantization parameter and this model has been applied to the reference software HM.

The structure of this paper is as follows: In Sect. 2, we briefly introduce the rate control technology in H.265/HEVC. In Sect. 3, we analyze and compare the rate-distortion performance of the R-λ model and the R-D model, then establish an optimization scheme for the bit allocation at the CTU level. Section 4 shows the experimental results of R-D performance. The conclusion is drawn in Sect. 5.

2 Review on the RC Scheme

2.1 Rate Control

As a part of the rate-distortion optimization, rate control can be described as the rate control method selects the coding parameters for each coding unit to minimize the distortion under the condition that the total number of coded bits is less than or equal to the target number of bits. The coding parameters include the weight of bit allocation, and the quantization parameters. It is formulated by

$$P^* = (p_1^*, \ldots, p_N^*) = \arg \min_{(p_1,\ldots,p_N)} \sum_{i=1}^{N} d_a \text{ s.t.} \sum_{i=1}^{N} r_i \leq R_t \tag{1}$$

where P^* is a set of optimal encoding parameters to satisfy the minimal distortion. N is the number of coding units which can be the number of GOP, frame or CTU; d_a and r_a are the distortion and the number of target bits for the ith CU, respectively; R_t is the total number of target bits under the constraint condition.

In general, the constraint issue can be converted by the Lagrange optimization method [5]. The Lagrange cost function is

$$J = \min \sum_{i=1}^{N} (d_i + \lambda r_i) \tag{2}$$

where λ is the slope of the R-D curve, and Eq. (2) can be expressed by

$$\lambda = -\frac{\partial d_i}{\partial r_i} \text{ and } \sum_{i=1}^{N} r_i = R \tag{3}$$

In the past decades, many R-D models [6–8] have been proposed, such as the quadratic model, the linear model, the logarithmic model and the exponential model. Li et al. proposed a hyperbolic model which performs better than the other models. It can be expressed by

$$D(R) = CR^{-K} \tag{4}$$

where C and K are the model parameters related to the characteristic of the picture. Therefore, the R-λ model can be rewritten as

$$\lambda = -\frac{\partial d_i}{\partial r_i} = C_i K_i R^{-K_i-1} = C_i K_i (bpp_i \cdot N_i)^{-K_i-1} = \alpha \cdot bpp_i^{\beta} \tag{5}$$

where $\alpha_i = C_i K_i \cdot N_i^{-K_i-1}$ and $\beta_i = -K_i - 1$ relate to the video content, and bpp_i indicates the number of bits per pixel. If target number of bits of a CTU or a frame is T and the number of pixels is N, the bpp is

$$bpp = \frac{T}{N} \tag{6}$$

2.2 Bit Allocation

The target bit allocation is one of the goals of the RC scheme. According to the target bitrate, the encoder allocates the target bits for different coding units using a grading strategy from GOP level to CTU level. Since the rate control cannot guarantee that the actual number of coded bits is exactly the same as that of the target bits, a data buffer is set up between the encoder and the channel. When the number of coded bits exceeds the capacity of the buffer, video delay is caused, and this is often referred as overflow. At GOP level, the number of target bits is determined by the buffer status and the number of frames in the current GOP. At frame level, the target bit is determined by the remaining bits in the current GOP and the target bpp level. At CTU (LCU) level, the number of target bits for each CTU is determined by

$$T_{CTU} = \frac{T_{pic} - Bit_{head} - Coded_{pic}}{\sum_{NotCodedCTUs} \omega_{CTU}} \times \omega_{CurCTU} \tag{7}$$

where the Bit_{head} is the estimated bits for all header information, including coding parameters such as coding mode, motion vector and quantization parameters. T_{pic} is the target bits of the current picture and $Coded_{pic}$ is the number of bits generated by the current picture. In K0103, ω is the weight for each CTU and it is measured by the predicted MAD, as calculated by Eq. (8)

$$\omega_{CTU} = MAD_{CTU}^2 \tag{8}$$

$$MAD_{CTU} = \frac{1}{N_{pixels}} \sum_i |pred_i - org_i| \tag{9}$$

After HM10.0, the bit allocation weight for each CTU is determined by

$$\omega_{CTU} = \left(\frac{\lambda_p}{\alpha_{CTU}}\right)^{\left(\frac{1}{\beta_{CTU}}\right)} \tag{10}$$

where λ_p is the estimated λ at the frame level. The mode parameters α_{CTU} and β_{CTU} are determined by the video content.

3 The Proposed R-D Model

As discussed above, the R-λ model is the key of the RC scheme, but the R-λ model is constructed to make the R-D model more accurate in calculating the quantitative parameters. In the encoding process, we can use the frame-level target *bpp* to obtain the estimated λ of the R-λ model and the estimated distortion of the R-D model. After the encoding, we can use the actual *bpp* to calculate the actual distortion and the actual λ. As the frame-level coding parameters affect the bit allocation at the CTU level, we propose a rate-distortion model to optimize the target bit allocation at the CTU level.

3.1 R-λ Model and R-D Model

Combining the hyperbolic models in Eq. (4) and in Eq. (5), the estimated distortion and the estimated λ can be obtained by

$$\begin{cases} D_{est,a} = C_a R_{est,a}^{-K_a} \\ \lambda_{est,a} = C_a K_a R_{est,a}^{-K_a-1} \end{cases} \tag{11}$$

where $D_{est,a}$ and $\lambda_{est,a}$ are the estimated distortion and the estimated λ of the ath frame. The actual distortion and the actual λ are calculated by

$$\begin{cases} D_{act,a} = C_a R_{act,a}^{-K_a} \\ \lambda_{act,a} = C_a K_a R_{act,a}^{-K_a-1} \end{cases} \tag{12}$$

In our experiments, the videos were processed by the HM14.0 reference software, with the lowdelay_P_main configuration. These two polyline figures are produced from coding the BasketballPass sequence.

Figure 1 shows the polylines of the estimated λ and the actual λ. Figure 2 shows the polylines of the estimated distortion and the actual distortion. Obviously, the difference between the two distortion polylines is smaller than that of the λ polylines. Therefore, the following section focuses on constructing an R-D model to optimize the CTU-level bit allocation.

3.2 Proposed R-D Model

In order to establish an accurate R-D model, we adjust the hyperbolic model in Eq. (4) by

$$d = C_i R_i^{-K_i} = C_i (bpp_i \cdot N_i)^{-K_i} = c_i \cdot bpp_i^{k_i} \tag{13}$$

where $c_i = C_i N_i^{-K_i}$ and $k_i = -K_i$,C_i, K_i are the model parameters at the CTU level. In the practical coding process, C_i and K_i of each CTU are different. According to Eqs. (3), (13) can be rewritten as

$$\sum_{i=1}^{N} r_i = \sum_{i=1}^{N} (bpp_i \cdot N_i) = \sum_{i=1}^{N} \left(\left(\frac{d}{c_i} \right)^{1/k_i} \cdot N_i \right) = R \tag{14}$$

As C_i and K_i are different, c_i and k_i are different from CTU to CTU as well. Similar to Eq. (10), the following equations can be used for bit allocation at the CTU level:

$$\begin{cases} r_i = R_{pic} \cdot \dfrac{\omega_i}{\sum_{j=1}^{N} \omega_j} \\ \omega_i = \left(\dfrac{d_p}{c_i} \right)^{\left(\frac{1}{k_i} \right)} \end{cases} \tag{15}$$

In order to solve the constraint problem, Eqs. (11) and (14) are combined to obtain the bit allocation at the CTU level

$$\sum_{i=1}^{N} r_i = \sum_{i=1}^{N} \left(\frac{d}{C_i} \right)^{-\frac{1}{K_i}} = \sum_{i=1}^{N} \left(\frac{C_i}{d} \right)^{a_i} = R \tag{16}$$

where $a_i = 1/K_i$. In order to make it easier to obtain d, we replace $(C_i/d)^{a_i}$ by $\tilde{r}_i(\tilde{d}/d)^{a_i}$, where $\tilde{d}((\tilde{d}/d) \to 1)$ is the estimated distortion, and $\tilde{r}_i = (C_i/\tilde{d})^{b_i}$ is the target bit for the ith CTU. Then Eq. (16) can be rewritten as

$$\sum_{i=1}^{N} \left(\frac{C_i}{d} \right)^{a_i} = \sum_{i=1}^{N} \tilde{r}_i \left(\frac{\tilde{d}}{d} \right)^{a_i} = R \tag{17}$$

We utilize the Taylor expansion to express $(\tilde{d}/d)^{a_i}$ as follows:

$$\left(\frac{\tilde{d}}{d} \right)^{a_i} = 1 + \frac{\ln(\frac{\tilde{d}}{d})}{1!} a_i + \ldots + \frac{\ln(\frac{\tilde{d}}{d})^n}{n!} a_i^n + \ldots \tag{18}$$

We retain the third-order items in the Taylor expansion, which is shown as:

$$\begin{aligned} \left(\frac{\tilde{d}}{d} \right)^{a_i} &\approx 1 + \frac{\ln(\frac{\tilde{d}}{d})}{1!} a_i + \frac{\ln(\frac{\tilde{d}}{d})^2}{2!} a_i^2 + \frac{\ln(\frac{\tilde{d}}{d})^3}{3!} a_i^3 \\ &= -\frac{a_i^3}{6} \ln^3 d + (\frac{a_i^2}{2} + \frac{a_i^3}{2} \ln \tilde{d}) \ln^2 d \\ &\quad - (a_i^2 \ln \tilde{d} + a_i + \frac{a_i^3}{2} \ln^2 \tilde{d}) \ln d \\ &\quad + (1 + a_i \ln \tilde{d} + \frac{a_i^2}{2} \ln^2 \tilde{d} + \frac{a_i^3}{6} \ln^3 \tilde{d}) \end{aligned} \tag{19}$$

Therefore the third-order Taylor approximation of Eq. (17) can be expressed as

$$\begin{aligned} R = \sum_{i=1}^{N} \tilde{r}_i \left(\frac{\tilde{d}}{d}\right)^{a_i} \approx & \underbrace{-\sum_{i=1}^{N} (\tilde{r}_i \frac{a_i^3}{6})}_{A} \ln^3 d \\ & \underbrace{+ \sum_{i=1}^{N} \tilde{r}_i (\frac{a_i^2}{2} + \frac{a_i^3}{2} \ln \tilde{d})}_{B} \ln^2 d \\ & \underbrace{- \sum_{i=1}^{N} \tilde{r}_i (a_i^2 \ln \tilde{d} + a_i + \frac{a_i^3}{2} \ln^2 \tilde{d})}_{C} \ln d \\ & \underbrace{+ \sum_{i=1}^{N} \tilde{r}_i (1 + a_i \ln \tilde{d} + \frac{a_i^2}{2} \ln^2 \tilde{d} + \frac{a_i^3}{6} \ln^3 \tilde{d})}_{D} \end{aligned} \tag{20}$$

To solve the third-order equation, the practical condition is $\Delta = F^2 - 4EG > 0$, where $E = B^2 - 3AC$, $F = BC - 9A(D - R)$, $G = C^2 - 3B(D - R)$. Then, we can obtain the unique d by

$$\begin{cases} d = e^{\frac{-B-(\sqrt[3]{Y_1} + \sqrt[3]{Y_2})}{3A}} \\ Y_{1,2} = BE + 3A(\frac{-F \pm \sqrt{F^2 - 4EG}}{2}) \end{cases} \tag{22}$$

Finally, we combined Eqs. (15) and (22) to obtain the bit weight of CTUs.

In this section, we found that estimate distortion model has a better similarity than estimate lambda model. We used the Taylor expansion formula to convert the constraint problem, and established an R-D model to optimize the weight of bit allocation.

4 Experimental Evaluation

To evaluate the performance of the proposed model, we implement our model on the reference software HM14.0, and compare it with HM14.0 in terms of bitrate error and PSNR.

The configuration file used is *encoder_lowdelay_P_main.cfg*. The rate control parameters are set as follows: RateControl: enabled; KeepHierarchicalBit: enabled; LCULevelRateControl: enabled; RCLCUSeparateModel: enabled; InitialQP: 0; RCForceIntraQP: disabled.

4.1 R-D Performance

The peak signal-to-noise ratio (PSNR) is the most common evaluation criterion to compare the quality changes of image. It is defined as

$$PSNR = 10\log_{10}\frac{f_{\max}^2}{MSE}$$

$$MSE = \frac{1}{W \times H}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left(p_{ori}(x,y) - p_{rec}(x,y)\right)^2 \tag{23}$$

where $p_{ori}(x,y)$ and $p_{rec}(x,y)$ represent the pixel value of the original image and the reconstructed image at the point (x,y), $f_{\max}$ indicates the peak value of the pixel.

4.2 Figures and Tables

In order to examine the effectiveness of our bit allocation scheme, we use line charts and tables to compare the performance more vividly.

The encoding results of the video Kimono1 and FourPeople are presented in Table 1. With the same coding environment and target bits, we encoded the videos by using HM14.0 and the proposed algorithm, respectively. From the data in the Table 1, some results are not ideal when encoding a video using different QP, such as Kimono1 in the case of QP = 37, FourPeople in the case of QP = 22. Therefore, in order to obtain a universal effect, we get their MPSNR and average bitrate error. Compared to HM14.0, the MPSNR increases of Kimono1 and FourPeople are 0.23 dB and 0.30 dB respectively. The average bitrate errors increase 0.07% and 0.16%, respectively. To illustrate the rate control performance more intuitively, we denoted them with line charts.

Table 1 Some results of HM14.0 and our proposed schemes

Kimono1	Target bit (kbps)	HM14.0		Proposed		Compare	
		Bitrate (kbps)	PSNR (dB)	Bitrate	PSNR	△kbps	△ PSNR
22	5482.135	5482.392	41.309	5484.802	41.396	0.05%	0.09
27	2472.606	2472.810	39.117	2473.605	39.402	0.04%	0.29
32	1194.99	1195.124	36.688	1195.495	36.971	0.04%	0.28
37	588.306	588.448	34.139	589.299	34.386	0.16%	0.25
Average						0.07%	0.23
Four people							
22	2729.193	2729.58	42.289	2736.465	42.396	0.27%	0.11
27	1007.426	1007.265	39.903	1008.9	40.334	0.15%	0.43
32	490.228	490.051	37.523	490.792	37.854	0.12%	0.33
37	259.783	259.711	35.007	260.047	35.327	0.10%	0.32
Average						0.16%	0.30

The Rate-Distortion curves of the two sequences are plotted in Figs. 3 and 4. They show that PSNR has gain at the same bitrate. The performance of our proposed scheme is better than RC scheme of HM14.0 both in video sequences of Kimono1 and FourPeople.

Table 2 shows the bitrate mismatch and the PSNR comparison between HM14.0 rate control algorithm and the proposed rate control algorithm. It can be observed that the PSNR is improved with less bitrate error.

Table 2 Comparison of all results

	Average △ kbps	Average △ PSNR
Kimonol	0.07%	0.23 dB
Four people	0.16%	0.30 dB
Blowing bubbles	0.54%	0.09 dB
Basketpass	0.12%	0.08 dB
BQMALL	0.38%	0.06 dB
Race horse	0.05%	0.07 dB
Party scene	0.03%	0.14 dB
Average	0.19%	0.14 dB

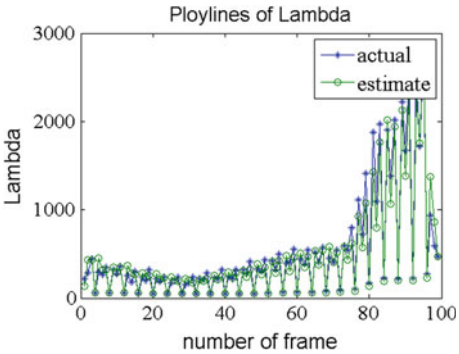


Fig. 1 Polyline of lambda

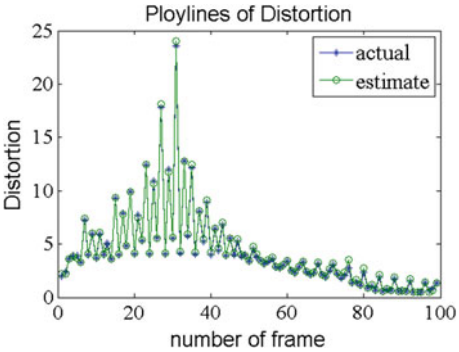


Fig. 2 Polyline of distortion

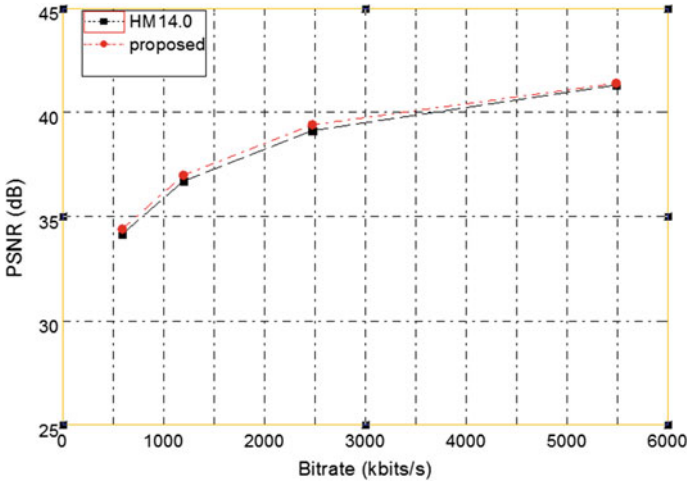


Fig. 3 Rate-distortion curve of Kimono1

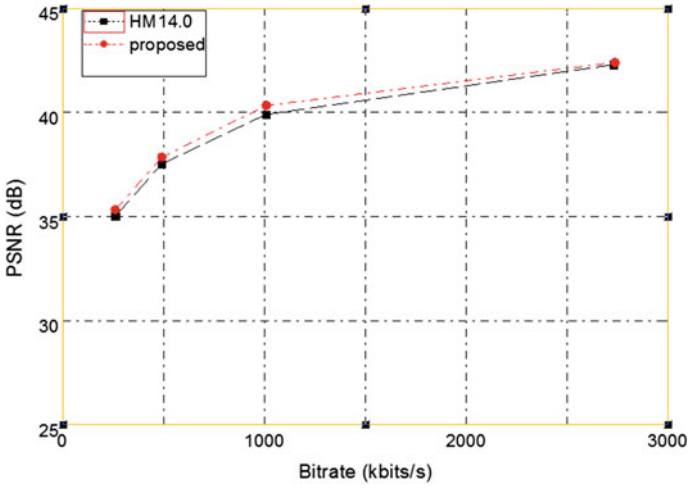


Fig. 4 Rate-distortion curve of FourPeople

5 Conclusion

In this paper, we proposed a novel R-D model for rate control, which defines a bit allocation scheme at the CTU level. Based on the R-D model, we determined the weight of bit allocation for each CTU in the current frame. In summary, our bit allocation scheme is able to obtain a better performance in terms of the R-D performance. The improvement which is averaged over these video sequences is an increase in PSNR of 0.14 dB with bitrate error of 0.19%.

Acknowledgment. This work has been supported by the National Natural Science Foundation of China (NSFC) under project No.61401123, the Fundamental Research Funds for the Central Universities under grant No. HIT.NSRIF.201617, and the Harbin Science and Technology Bureau under project No.2014RFQXJ166.

References

1. Li, B., Li, H., Li, L.: Adaptive Bit Allocation for R-Lambda Model Rate Control in HM document JCTVC-M0036 (2013)
2. ITU-T VCEG, Joint Call for Proposals on Video Compression Technology, document ITU-T VCEG-AM91, and ISO/IEC MPE N11113, Kyoto, Japan (2010)
3. Ahmad, I., Luo, J.: On using game theory to optimize the rate control in video coding. IEEE Trans. Circuits Syst. Video Technol. **16**, 209–219 (2006)
4. Li, B., Li, H., Li, L., Zhang, J.: Rate control by R-lambda model for HEVC. ITU-T/ISO/IEC JCT-VC Document JCTVC-K0103, QCT, Shanghai (2012)
5. Choi, H., Yoo, J., Nam, J., Sim, D., Bajic I.V.: Pixel-wise unified rate-quantization model for multi-level rate control. IEEE J. Sel. Top. Signal Process. **7**(6), 1112–1123 (2013)
6. Lin, L.-J., Ortega, A.: Bit-rate control using piecewise approximated rate-distortion characteristics. IEEE Trans. Circuits Syst. Video Technol. **8**(4), 446–459 (1998)
7. Wang, M., van der Schaar, M.: Operational rate-distortion modeling for wavelet video coders. IEEE Trans. Signal Process. **54**(9), 3505–3517 (2006)
8. Tu, Y.-K., Yang, J.-F., Sun, M.-T.: Rate-distortion modeling for efficient H.264/AVC encoding. IEEE Trans. Circuits Syst. Video Technol. **17**(5), 530–543 (2007)

On Equivalence of GSVD-SLNR Precoding and JD-SLNR Precoding

Zuoliang Yin[1(✉)] and Zhian Deng[2]

[1] School of Electronic Information, Qingdao University, Qingdao 266071, China
yinzuoliang@qdu.edu.cn
[2] College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China
dengzhian@hrbeu.edu.cn

Abstract. GSVD-SLNR precoding and JD-SLNR precoding are two remarkable linear precoding schemes in downlink MU-MIMO systems. Both of them outperform the original SLNR precoding with considerable gains in terms of bit error rate. However, the relationship of such two schemes has not been investigated yet. In this letter, the equivalence of such two schemes is proved theoretically, and simulation results are given to support the view. The equivalence indicates that the performance improvement of GSVD-SLNR scheme over an original SLNR precoding attributes to the reduction of the gap between the SINRs of multiple data streams.

Keywords: Multiple-input Multiple-output (MIMO) systems · Signal-to-leakage-plus-noise ratio (SLNR) · Precoding · Generalised singular value decomposition (GSVD) · Joint decomposition (JD)

1 Introduction

Recently, signal-to-leakage-plus-noise ratio (SLNR) based precoding attracts considerable attention in both single-cell and multicell multiple-input multiple-output (MIMO) downlink due to its potential for effective co-channel interference (CCI) mitigation [1]. The original SLNR precoding is proposed by Targhat and Zhang [2,3] for one stream per user, and Sadek et al. [4] expands it to a multi-stream scenario. SLNR-based precoding schemes are also considered in massive MIMO systems, full-duplex MU-MIMO Systems and heterogeneous networks [5–8]. The equivalence of SLNR precoding and regularised block diagonalisation (RBD) is demonstrated [9,10]. To further improve bit error rate (BER) performance of SLNR-based precoding, Park et al. [11] and Cheng et al. [12] proposed two improved SLNR-based precoding schemes recently. Park proposed a precoding scheme using generalised singular value decomposition (GSVD), referred to as the GSVD-SLNR precoding. Cheng proposed one based on joint diagonalisation (JD), referred to as the JD-SLNR precoding. Both of them outperform

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 302–307, 2019.
https://doi.org/10.1007/978-981-13-6264-4_36

the original SLNR precoding with sizable gains. Such two schemes are proposed based on different motivations and appear to be different. However, the relationship of these two schemes has not been shown yet.

In this study, we prove that the GSVD-SLNR precoding is equivalent to JD-SLNR precoding and they have a same BER performance. The proof is derived theoretically and verified by simulation results. Based on such equivalence, it can be obtained that the reason the GSVD-SLNR precoding outperforms the original SLNR-based precoding is the reduction of gap between the SINRs of multiple data streams.

Notations: $\mathrm{Tr}(\cdot)$, $\mathrm{E}(\cdot)$, $(\cdot)^H$, and $(\cdot)^{-1}$ denote trace, expectation, conjugate transpose, and inverse, respectively. $\mathbb{C}^{M\times N}$ is the set of $M \times N$ matrices in complex field. $\mathbf{I}_L$ represents the $L \times L$ identify matrix. $\mathbf{A} \triangleq \mathbf{B}$ denotes that $\mathbf{A}$ is defined as $\mathbf{B}$.

2 System Model

We consider a MU-MIMO downlink system with N transmitting antennas and M receiving antennas at each of K active users. The transmit signal $\mathbf{s}_k \in \mathbb{C}^{L\times 1}$ for user k $(k = 1, \ldots, K)$ is first passed the corresponding precoding matrix $\mathbf{W}_k \in \mathbb{C}^{N\times L}$, where $\mathbf{s}_k$ and $\mathbf{W}_k$ satisfy power limitation $\mathrm{E}(\mathbf{s}_k\mathbf{s}_k^H) = \mathbf{I}_L$ and $\mathrm{Tr}(\mathbf{W}_k\mathbf{W}_k^H) = L$, respectively. The output signal is then transmitted by N antennas. Let $\mathbf{H}_k \in \mathbb{C}^{M\times N}$ denote the channel for user k, and its elements are modelled as independent and identically distributed complex Gaussian variables with zero-mean and unit-variance. Define $\bar{\mathbf{H}}_k = [\mathbf{H}_1^H, \mathbf{H}_2^H, \ldots, \mathbf{H}_{k-1}^H, \mathbf{H}_{k+1}^H, \ldots, \mathbf{H}_K^H]^H$. The received signal for user k can be expressed as

$$\mathbf{r}_k = \mathbf{H}_k\mathbf{W}_k\mathbf{s}_k + \mathbf{H}_k \sum_{i=1, i\neq k}^{K} \mathbf{W}_i\mathbf{s}_i + \mathbf{z}_k, \tag{1}$$

where $\mathbf{z}_k$ denotes the additive complex Gaussian noise vector, and its elements are zero-mean random variables with variance σ^2.

3 Equivalence of GSVD-SLNR and JD-SLNR Precoding

3.1 The GSVD-SLNR Precoding

We first review the GSVD-SLNR precoding briefly. GSVD-SLNR precoding is based on the following theorem:

Theorem 1 [11]: Given that two matrices $\mathbf{H}_b \in \mathbb{C}^{m\times n}$ and $\mathbf{H}_\omega \in \mathbb{C}^{p\times n}$ have the same number of columns, then there exists unitary matrices $\mathbf{U} \in \mathbb{C}^{m\times m}$, $\mathbf{V} \in \mathbb{C}^{p\times p}$, and a nonsingular matrix $\mathbf{X} \in \mathbb{C}^{n\times n}$ such that

$$\mathbf{U}^H\mathbf{H}_b\mathbf{X} = [\mathbf{\Sigma}_b, \mathbf{0}], \tag{2}$$

$$\mathbf{V}^H\mathbf{H}_\omega\mathbf{X} = [\mathbf{\Sigma}_\omega, \mathbf{0}], \tag{3}$$

where $\mathbf{\Sigma}_b \in \mathbb{C}^{m\times t}$, $\mathbf{\Sigma}_\omega \in \mathbb{C}^{p\times t}$, $t = \text{rank}([\mathbf{H}_b^H, \mathbf{H}_\omega^H]^H)$, and $\mathbf{\Sigma}_b^H\mathbf{\Sigma}_b + \mathbf{\Sigma}_\omega^H\mathbf{\Sigma}_\omega = \mathbf{I}_t$.

By substituting $\mathbf{H}_k$ and $[\bar{\mathbf{H}}_k^H, \sqrt{M\sigma}\mathbf{I}_N]^H$ for $\mathbf{H}_b$ and $\mathbf{H}_\omega$, respectively, the precoding weights are shown to be the first L columns of $\mathbf{X}$, i.e.

$$\mathbf{W}_k^{\text{GSVD}} = \mathbf{X}\,[\mathbf{I}_L, \mathbf{0}]^H\,. \tag{4}$$

3.2 The JD-SLNR Precoding

In [12], Cheng proposed a SLNR precoding based on joint decomposition of two Hermitian matrices. For analytical purpose, we briefly summarise the generating process of such precoding scheme.

Define $\mathbf{A}_k \triangleq \mathbf{H}_k^H\mathbf{H}_k$, $\mathbf{B}_k \triangleq M\sigma^2\mathbf{I}_N + \bar{\mathbf{H}}_k^H\bar{\mathbf{H}}_k$, and $\mathbf{C}_k \triangleq \mathbf{A}_k + \mathbf{B}_k$. There exists a nonsingular matrix $\mathbf{Q}_k \in \mathbb{C}^{N\times N}$ such that

$$\mathbf{Q}_k^H\mathbf{C}_k\mathbf{Q}_k = \mathbf{Q}_k^H(\mathbf{A}_k + \mathbf{B}_k)\mathbf{Q}_k = \mathbf{I}_N. \tag{5}$$

Define $\mathbf{A}'_k \triangleq \mathbf{Q}_k^H\mathbf{A}_k\mathbf{Q}_k$ and $\mathbf{B}'_k \triangleq \mathbf{Q}_k^H\mathbf{B}_k\mathbf{Q}_k$, then $\mathbf{A}'_k$ and $\mathbf{B}'_k$ have the same eigenvectors. The eigen-decomposition (ED) of $\mathbf{A}'_k$ and $\mathbf{B}'_k$ are

$$\mathbf{A}'_k = \mathbf{G}_k\mathbf{\Lambda}_b\mathbf{G}_k^H, \tag{6}$$

$$\mathbf{B}'_k = \mathbf{G}_k\mathbf{\Lambda}_\omega\mathbf{G}_k^H. \tag{7}$$

where $\mathbf{\Lambda}_b + \mathbf{\Lambda}_\omega = \mathbf{I}_N$. Let $\mathbf{T} = \mathbf{Q}_k\mathbf{G}_k$, then the precoding weights are the first L columns of $\mathbf{T}$, i.e.

$$\mathbf{W}_k^{\text{JD}} = \mathbf{T}\,[\mathbf{I}_L, \mathbf{0}]^H\,. \tag{8}$$

3.3 Equivalence of the Two Precoding Schemes

In this section, we prove that the GSVD-SLNR precoding is equivalent to the JD-SLNR. Specifically, we have the following theorem:

Theorem 2: The GSVD-SLNR precoding is equivalent to the JD-SLNR precoding, i.e. $\mathbf{W}_k^{\text{GSVD}} = \mathbf{W}_k^{\text{JD}}$, with $\mathbf{\Lambda}_b = \mathbf{\Sigma}_b^H\mathbf{\Sigma}_b$, $\mathbf{\Lambda}_\omega = \mathbf{\Sigma}_\omega^H\mathbf{\Sigma}_\omega$.

Proof. (a) GSVD-SLNR⇒JD-SLNR.

Set $\mathbf{H}_b = \mathbf{H}_k$ and $\mathbf{H}_\omega = \left[\bar{\mathbf{H}}_k^H, \sqrt{M\sigma}\mathbf{I}_N\right]^H$. Based on the assumption that the elements of $\mathbf{H}_k$ and $\bar{\mathbf{H}}_k$ are independent and identically distributed, $[\mathbf{H}_b^H, \mathbf{H}_\omega^H]^H$ has a full column rank, i.e. $t = \text{rank}([\mathbf{H}_b^H, \mathbf{H}_\omega^H]^H) = N$, and (2), (3) can be written as

$$\mathbf{U}^H\mathbf{H}_b\mathbf{X} = \mathbf{\Sigma}_b, \tag{9}$$

$$\mathbf{V}^H\mathbf{H}_\omega\mathbf{X} = \mathbf{\Sigma}_\omega. \tag{10}$$

It follows that

$$\mathbf{A}_k = \mathbf{H}_b^H\mathbf{H}_b = \mathbf{X}^{-H}\mathbf{\Sigma}_b^H\mathbf{\Sigma}_b\mathbf{X}^{-1}, \tag{11}$$

$$\mathbf{B}_k = \mathbf{H}_\omega^H\mathbf{H}_\omega = \mathbf{X}^{-H}\mathbf{\Sigma}_\omega^H\mathbf{\Sigma}_\omega\mathbf{X}^{-1}. \tag{12}$$

Since $\mathbf{\Sigma}_b^H\mathbf{\Sigma}_b + \mathbf{\Sigma}_\omega^H\mathbf{\Sigma}_\omega = \mathbf{I}_N$, we have

$$\mathbf{X}^H(\mathbf{A}_k + \mathbf{B}_k)\mathbf{X} = \mathbf{I}_N. \tag{13}$$

Let $\mathbf{G}_k \in \mathbb{C}^{N\times N}$ be any unitary matrix, it follows that

$$(\mathbf{X}\mathbf{G}_k^H)^H(\mathbf{A}_k + \mathbf{B}_k)(\mathbf{X}\mathbf{G}_k^H) = \mathbf{I}_N. \tag{14}$$

It can be observed that (5) holds if we set $\mathbf{Q}_k = \mathbf{X}\mathbf{G}_k^H$, i.e. $\mathbf{X} = \mathbf{Q}_k\mathbf{G}_k$. From (6), (7) we obtain

$$\mathbf{G}_k^H\mathbf{A}_k'\mathbf{G}_k = \mathbf{G}_k^H\mathbf{Q}_k^H\mathbf{A}_k\mathbf{Q}_k\mathbf{G}_k = \mathbf{X}^H\mathbf{A}_k\mathbf{X} = \mathbf{\Lambda}_b, \tag{15}$$

$$\mathbf{G}_k^H\mathbf{B}_k'\mathbf{G}_k = \mathbf{G}_k^H\mathbf{Q}_k^H\mathbf{B}_k\mathbf{Q}_k\mathbf{G}_k = \mathbf{X}^H\mathbf{B}_k\mathbf{X} = \mathbf{\Lambda}_\omega. \tag{16}$$

Compare (15), (16) with (11), (12), we have $\mathbf{\Lambda}_b = \mathbf{\Sigma}_b^H\mathbf{\Sigma}_b$, $\mathbf{\Lambda}_\omega = \mathbf{\Sigma}_\omega^H\mathbf{\Sigma}_\omega$.

(b) JD-SLNR⇒GSVD-SLNR.

We recall the definition that $\mathbf{A}_k' = \mathbf{Q}_k^H\mathbf{A}_k\mathbf{Q}_k = \mathbf{Q}_k^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{Q}_k$. It follows from (6) that

$$\mathbf{Q}_k^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{Q}_k = \mathbf{G}_k\mathbf{\Lambda}_b\mathbf{G}_k^H. \tag{17}$$

In general, there is no unique decomposition on (17) because $\mathbf{H}_k^H\mathbf{H}_k = \mathbf{H}_k^H\mathbf{U}\mathbf{U}^H\mathbf{H}_k$ holds for any unitary matrix $\mathbf{U}$. We have

$$\mathbf{Q}_k^H\mathbf{H}_k^H\mathbf{U}\mathbf{U}^H\mathbf{H}_k\mathbf{Q}_k = \mathbf{G}_k\mathbf{\Lambda}_b\mathbf{G}_k^H, \tag{18}$$

$$\mathbf{U}^H\mathbf{H}_k\mathbf{Q}_k = \bar{\mathbf{\Sigma}}_b\mathbf{G}_k^H, \tag{19}$$

$$\mathbf{U}^H\mathbf{H}_k\mathbf{Q}_k\mathbf{G}_k = \bar{\mathbf{\Sigma}}_b, \tag{20}$$

where $\bar{\mathbf{\Sigma}}_b \in \mathbb{C}^{M\times N}$ and $\mathbf{\Lambda}_b = \bar{\mathbf{\Sigma}}_b^H\bar{\mathbf{\Sigma}}_b$. If we define $\mathbf{X} \triangleq \mathbf{Q}_k\mathbf{G}_k$, then

$$\mathbf{U}^H\mathbf{H}_k\mathbf{X} = \bar{\mathbf{\Sigma}}_b. \tag{21}$$

Similarly, we can obtain that

$$\mathbf{V}^H\left[\bar{\mathbf{H}}_k^H, \sqrt{M}\sigma\mathbf{I}_N\right]^H\mathbf{X} = \bar{\mathbf{\Sigma}}_\omega, \tag{22}$$

where $\mathbf{V}$ is any unitary matrix, $\bar{\mathbf{\Sigma}}_\omega \in \mathbb{C}^{(N+(K-1)M)\times N}$, $\mathbf{\Lambda}_\omega = \bar{\mathbf{\Sigma}}_\omega^H\bar{\mathbf{\Sigma}}_\omega$. Compare (21), (22) with (9), (10), it is easy to see that $\bar{\mathbf{\Sigma}}_b = \mathbf{\Sigma}_b$ and $\bar{\mathbf{\Sigma}}_\omega = \mathbf{\Sigma}_\omega$, so that $\mathbf{\Lambda}_b = \mathbf{\Sigma}_b^H\mathbf{\Sigma}_b$, $\mathbf{\Lambda}_\omega = \mathbf{\Sigma}_\omega^H\mathbf{\Sigma}_\omega$. Since $\mathbf{\Lambda}_b + \mathbf{\Lambda}_\omega = \mathbf{I}_N$, we have $\mathbf{\Sigma}_b^H\mathbf{\Sigma}_b + \mathbf{\Sigma}_\omega^H\mathbf{\Sigma}_\omega = \mathbf{I}_N$, which satisfies the constraint of GSVD-SLNR.

Lastly, we have to prove $\mathbf{X}$ is nonsingular. This is obvious because $\mathbf{Q}_k$ and $\mathbf{G}_k$ are both nonsingular. □

Theorem 2 shows that the GSVD-SLNR precoding and JD-SLNR precoding are equivalent so that they are expected to have the same bit error rate (BER) performance. Moreover, it can be obtained from (15), (16) that the GSVD-SLNR precoding jointly diagonalisates the channel covariance matrix of the desired user and the leakage channel-plus-noise covariance matrix. Therefore, the reason GSVD-SLNR precoding outperforms the original SLNR-based precoding is that the GSVD-SLNR precoding reduces the gap between the effective SINRs of multiple data streams, which has been demonstrated in detail [12].

4 Simulation Results

Figure 1 compares the uncoded BER performance of GSVD-SLNR precoding [11] with JD-SLNR precoding [12] in a multi-user MIMO system downlink. The channel is modelled to be flat fading and is known at the transmitter. QPSK modulation with Gray mapping is assumed. In Fig. 1, BER curves are plotted versus E_b/N_0, i.e. $1/\sigma^2$. It is observed that the BER performance of GSVD-SLNR precoding is extremely close to the JD-SLNR one, which verifies the equivalence numerically to a certain extent.

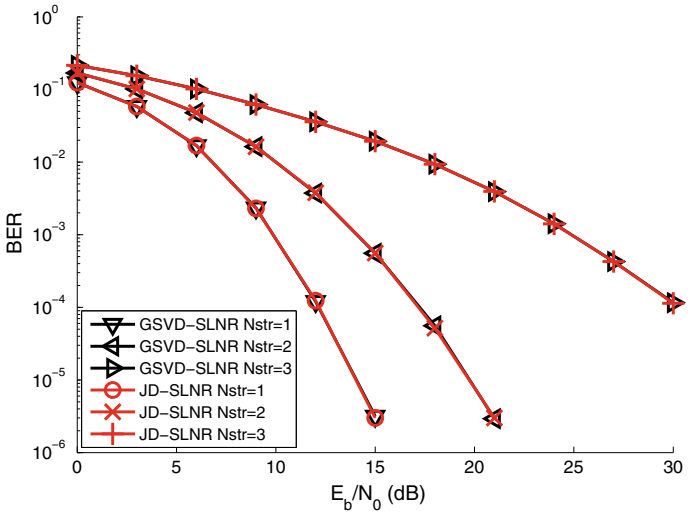


Fig. 1. Performance comparison of the GSVD-SLNR precoding with the JD-SLNR precoding in terms of BER. $N = 10$, $M = 3$, $K = 3$, Nstr denotes the number of data streams for each user.

5 Conclusion

In this letter, we proved that the GSVD-SLNR precoding and the JD-SLNR precoding are equivalent. This result indicates their close BER performance, which is also demonstrated by the simulation results. In addition, the equivalence explains the reason why the GSVD-SLNR precoding outperforms the original SLNR-based precoding, that is, the reduction of gap between SINRs of multiple data streams.

References

1. Fatema, N., Hua, G., Xiang, Y., Peng, D., Natgunanathan, I.: Massive MIMO linear precoding: a survey. IEEE Syst. J. **PP**(99), 1–12 (2017)
2. Targhat, A., Sadek, M., Sayed, A.H.: A multi user beamforming scheme for downlink MIMO channels based on maximising signal-to-leakage ratios. In: Proceedings of IEEE ICASSP, vol. 3, pp. iii/1129–iii/1132 (2005)
3. Zhang, J., Wu, Y., Xu, M., Wang, J.: Linear transmitter precoding design for downlink of multiuser MIMO systems. Electron. Lett. **41**(14), 811–813 (2005)
4. Sadek, M., Tarighat, A., Sayed, A.H.: A leakage-based precoding scheme for downlink multi-user MIMO channels. IEEE Trans. Wirel. Commun. **6**(5), 1711–1721 (2007)
5. Wang, J., Zhang, G., Lv, G., Li, G.: Downlink SLNR based precoding and power control in massive MIMO systems with time-shifted pilots. In: IEEE International Conference on Communications Workshops (ICC), pp. 200–206 (2016)
6. Tran, T.-X., Teh, K.-C.: Spectral and energy efficiency analysis for SLNR precoding in massive MIMO systems with imperfect CSI. IEEE Trans. Wirel. Commun. 1–11 (2018) (Early Access)
7. Berhane, T., Meng, W.-X., Chen, L., Jobir, G., Li, C.: SLNR-based precoding for single cell full-duplex MU-MIMO systems. IEEE Trans. Veh. Technol. **66**(9), 7877–7887 (2017)
8. Boukhedimi, I., Kammoun, A., Alouini, M.: Coordinated SLNR based precoding in large-scale heterogeneous. IEEE J. Sel. Top. Signal Process. **66**(9), 7877–7887 (2017)
9. Wang, K., Zhang, X.: On equivalence of SLNR-based precoding and RBD precoding. Electron. Lett. **48**(11), 662–663 (2012)
10. Patcharamaneepakorn, P., Doufexi, A., Armour, X.: Equivalent expressions and performance analysis of SLNR precoding schemes: a generalisation to multi-antenna receivers. IEEE Commun. Lett. **17**(6), 1196–1199 (2013)
11. Park, J., Chun, J., Park, H.: Generalised singular value decomposition based algorithm for multi-user multiple-input multiple-output linear precoding and antenna selection. IET Commun. **4**(16), 1899–1907 (2010)
12. Cheng, P., Tao, M., Zhang, W.: A new SLNR-based linear precoding for downlink multi-user multi-stream MIMO systems. IEEE Commun. Lett. **14**(11), 1008–1010 (2010)

A Research on Transmission Performance of MIMO-OFDM Wireless Channel

Pengxiang Li[1,2], Ruiqing Zhang[1,2], and Hui Li[1,2(✉)]

[1] College of Information Science and Technology, Hainan University, Haikou, China
nzlpx@foxmail.com, {iezhangruiqing,hitlihuilll2}@163.com

[2] Marine Communication and Network Engineering Technology Research Center of Hainan Province, Hainan University, Haikou, China

Abstract. Environment of mobile communication is very complex since the radio wave will not only increase loss with the propagation distance, but also produce multipath effect by the terrain and the obstacles. When the channel adopts MIMO technology, channel capacity is improved without increasing bandwidth and transmitted power. The OFDM technology can effectively deal with the interference between the signal waveform, and it is suitable for the high-speed data transmission multipath environment and fading channel. In this paper, through theoretical research on MIMO-OFDM techniques, we modeled and simulated wireless communication channel to better solve the high-quality transmission problem.

Keywords: IMO · OFDM · Space division multiplexing · Channel capacity

1 Introduction

At present, the research of Multiple-Input and Multiple-Output-Orthogonal Frequency Division Multiplexing (MIMO-OFDM) has entered the accelerated orbit, and the matured commercial products have been pushed to the market. A company of America Airgo Networks Inc. [1] was known as proposer of MIMO-OFDM technology. The development team has provided a number of IEEE 802.11n drafts using MIMO-OFDM technology, making the IEEE 802.11n draft standard to the world's first user of MIMO-OFDM technology standard, whose goal is to provide up to 54 Mbit/s data transfer rate. NTT DoCoMo was tested in the field using OFDM technology in May 2003 [2], and achieved the effective transmission of data rates in the 100 MHz bandwidth, and created its "peak rate" record later through a variety of technologies.

The world communication organization demonstrated the great ability of MIMO-OFDM technology to improve the transmission rate, and established the confidence in the application of MIMO-OFDM technology to commercial networks. At the same time, Intel and Texas Instruments and other chip manufacturers as well as Samsung, Nortel, and other telecommunication equipment manufacturers choose MIMO-OFDM technology based on the development of Worldwide Interoperability for Microwave Access (WiMax) technology. At present, the research of MIMO-OFDM network

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 308–316, 2019.
https://doi.org/10.1007/978-981-13-6264-4_37

access technology has been extended from the indoor wireless local area network to the wider metropolitan area network.

2 MIMO Channel Model

The main goal of developing next-generation wireless communication systems is to increase link throughput and network capacity. Significant improvements in throughput can be achieved when multiple antennas are applied at the transmitter and receiver, especially in a rich scattering environment, which has been used in flat fading [3] and in the wireless communication road of frequency-selective fading channels.

In wireless communication, it uses multiple antennas to transmit and receive, which transmits the signal through multiple spatial paths to transmit segmentation of the total transmit power without consuming additional radio frequency, thereby increasing the overall spectral efficiency. In order to achieve long-distance wireless communications, it needs to use time-space signal processing and spatial diversity of the combination of technology. Time-space signal processing is used to implement space-time coding at the sending port, including Space-Time Block Codes (STBC) [4], Space-Time Trellis Code (STTC) [5], and Bell Labs Layered Space-Time (BLAST) [6].

2.1 Summary of Space-Time Coding

Space-time signal processing, in other words, is space-time coding technology; the current study is more stratified space-time trellis coding and space-time block coding. The space-time block coding was first proposed by Alamouti [7], it is specifically applied to two transmit antennas orthogonal space-time code, and later extended to three or more antenna situation. Figure 1 is the Alamouti send coding simple diagram.

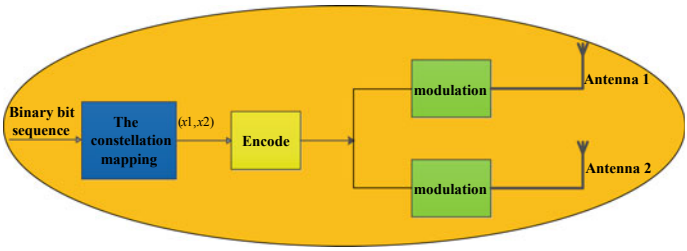


Fig. 1. Alamouti send diversity time-space coding scheme

The Alamouti encoded signal is transmitted from two transmit antennas via two symbol periods. In this example, the binary information bits sent by the source are the first proceeded constellation mapping [8].

The serial binary information sequence is input by serial-to-parallel conversion into $m = \log_2 M$ parallel data stream, where M is the number of constellation points of the constellation diagram, the each road data rate is R/m, R is the serial input code data rate. Each m bit corresponds to a constellation point on the constellation diagram, such as

Binary Phase Shift Keying (BPSK) modulation; each bit corresponds to a constellation point, such as Quadrature Phase Shift Keying (QPSK) modulation; each two bit corresponds to a constellation point, such as 16 Quadrature Amplitude Modulation (16QAM) modulation, every four bit corresponds to a constellation point. In this simulation, signal-to-noise ratio is 15 dB (SNR = 15 dB) in Figs. 2, 3, and 4.

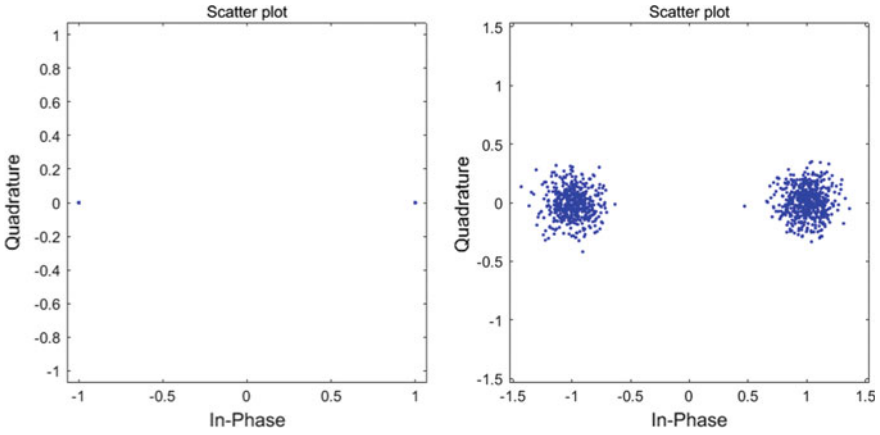


Fig. 2. BPSK modulation constellation and BPSK noise scatter

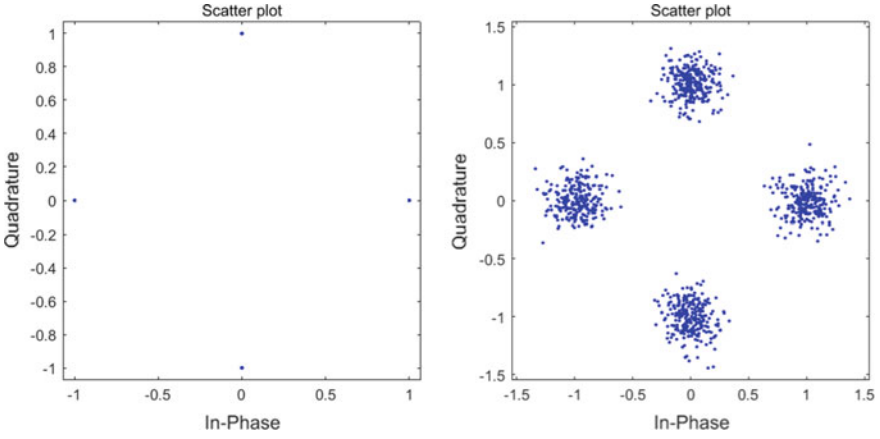


Fig. 3. QPSK modulation constellation and QPSK noise scatter

Here, we are assuming a quaternary modulation constellation with $m = \log_2 M = 2$. The binary information bits from the source are divided into two groups (in this case, it

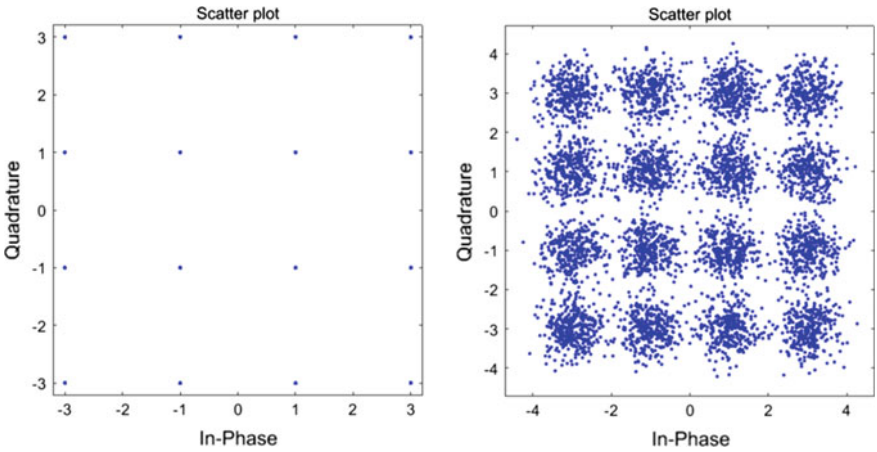


Fig. 4. 16 QAM modulation constellation and 16 QAM noise scatter

is x_1, x_2), and we get the two sets of bits from constellation mapping to obtain two modulation symbols x_1, x_2. Put these two symbols into the encoder and their encode form as follows:

$$\begin{bmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{bmatrix}. \tag{1}$$

The sent signals on the two transmit antennas satisfy the quadrature characteristics. Consider that there are two transmit antennas and one receiving antenna. Assume that the receiver can fully and accurately estimate the channel fading coefficients h_1 and h_2, use the maximum likelihood estimate at the receiver to find a pair of symbols $\left(\hat{x_1}, \hat{x_2}\right)$ from the constellation diagram, which the final receiver considers to be the sent symbol by the sender. The decision formula can be expressed as

$$\begin{aligned} \hat{x_1} &= \arg\min\left(|h_1|^2 + |h_2|^2 - 1\right)\left|\hat{x_1}\right|^2 + d^2\left(\tilde{x_1}, \hat{x_1}\right), \\ \hat{x_2} &= \arg\min\left(|h_1|^2 + |h_2|^2 - 1\right)\left|\hat{x_2}\right|^2 + d^2\left(\tilde{x_2}, \hat{x_2}\right), \end{aligned} \tag{2}$$

where the $\tilde{x_1}$ and $\tilde{x_2}$ signals are obtained by combining the channel fading coefficient and the received signal. In MIMO, the encoding and transmission schemes of the sender are the same as Single-Input Single-Output (SISO). Only the processing at the receiving end becomes complicated, it is necessary to combine and dispose of the signals received on the different receiving antennas. The decision metric under the multi-receive antenna can be obtained by linearly combining obtained from

the received signals on each pair of receiving antennas. The decision formula is as follows:

$$\hat{x_1} = \arg\min\left\{\left[\sum_{j=1}^{nr}\left(\left|h_{j,1}\right|^2 + \left|h_{j,2}\right|^2\right) - 1\right]\left|\hat{x_1}\right|^2 + d^2\left(\tilde{x_1}, \hat{x_1}\right)\right\},$$

$$\hat{x_2} = \arg\min\left\{\left[\sum_{j=1}^{nr}\left(\left|h_{j,1}\right|^2 + \left|h_{j,2}\right|^2\right) - 1\right]\left|\hat{x_2}\right|^2 + d^2\left(\tilde{x_2}, \hat{x_2}\right)\right\}. \quad (3)$$

2.2 Antenna Diversity Technology

For the Additive White Gaussian Noise (AWGN) channel, when the SNR is large, the slope of the Bit Error Rate (BER) curve approaches infinity. The relationship between BER and SNR (in E_b/N_0) is shown in Fig. 5. In other words, when the SNR increases, the channel exhibits a waterfall form BER performance. For Rayleigh fading channels, the corresponding slope of the curve is linear in logarithmic–logarithmic coordinates. It shows that even in the case of high signal-to-noise ratio, transmission performance through Rayleigh fading channel will be significantly reduced.

It can be seen that the MIMO model has a space-time encoder with multiple antennas, and its system model is consistent with the MIMO system theory described above. In general, $n_t > n_r$, because the number of antennas supported at mobile terminals is always less than at the base station. The reception vector is able to express $y = Hx + n$, the received signal is the channel fading coefficient transmission signal plus the noise of the receiving end. The capacity of the MIMO system can be expressed by the channel capacity (information transmission rate) of the Shannon formula, which is

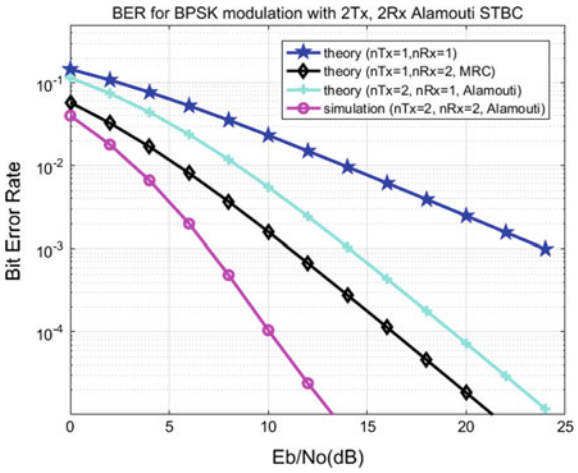


Fig. 5. BER for BPSK modulation with Alamouti (2 × 2) STBC

$$C = B\log_2(1 + S/N). \tag{4}$$

3 OFDM Modulation

One of the most effective means to overcome the frequency-selective fading is the multi-carrier transmission technique. A high-speed data flow is decomposed into a number of low-speed sub-data streams. Each sub-data stream is modulated into a sub-channel, which constitutes a number of parallel tuned signals, the final merger after transmission.

Now, OFDM modulation data transmission has N_s subcarrier [9]. The modulation rate is R_s/N_s. And, R_s is a single carrier rate. When the bits are generated, the OFDM code words and symbols can be obtained by modulation and coding. First, time domain transformation is performed using Inverse Fast Fourier Transform (IFFT). Second, the obtained signals are added by guard interval or cyclic prefix (CP). OFDM complex symbols S_1, S_2, …, S_i, and then through the serial-to-parallel conversion restructuring length of the block N_s, the transmission time is equal to $N_s \times T_s$. The $[iT_s, (i + 1)T_s]$ OFDM signal can be expressed as

$$\mathrm{x}l(t) = \sum_{k=\frac{Ns}{2}}^{\frac{Ns}{2}-l} Sl, kg(t - ((i+1)Ts))\exp(j2\,\pi fkt). \tag{5}$$

The above equation represents the l^{th} OFDM symbol. The received signal may be expressed as

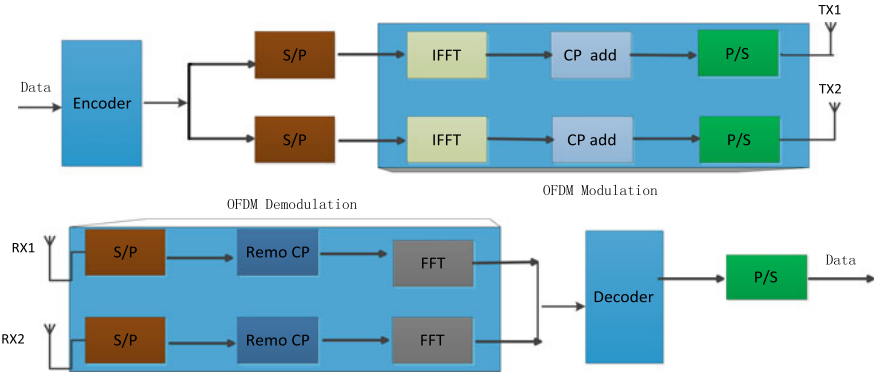


Fig. 6. Transmission model for OFDM systems

$$r(t) = \sum_{l=1}^{\infty} \sum_{k=-\frac{Ns}{2}}^{\frac{Ns}{2}-l} Sl, kg(t - ((i+1)Ts)) \exp(j2\,\pi fkt) + nk(t), \tag{6}$$

where $n_k(t)$ is the Gaussian white noise at the k^{th} subcarrier.

The above figure is 64 subcarrier channels, each channel has 4000 bits, using BPSK digital modulation mapping to achieve, this is a simple example of OFDM modulation, in fact, a complete OFDM system also includes interleaving, de-interleaving, and windowing section, which are omitted in Fig. 6.

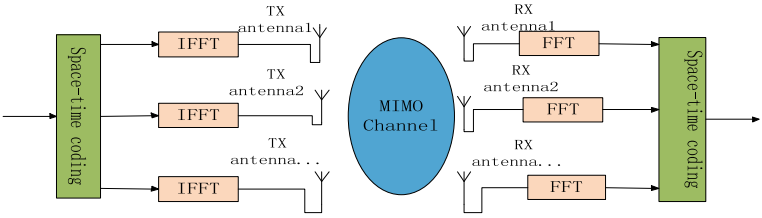


Fig. 7. STBC-MIMO-OFDM system block diagram

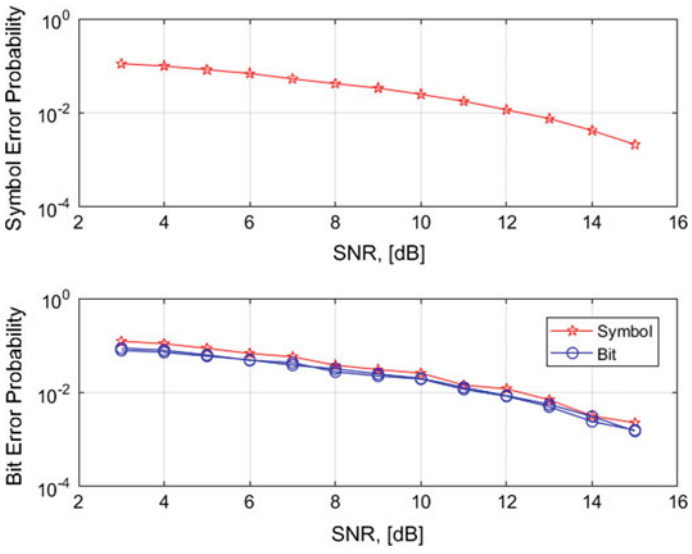


Fig. 8. STBC-MIMO-OFDM transmission characteristics

4 MIMO-OFDM System Combination

In order to achieve the expansion of MIMO and OFDM needs, the baseband signal processing needs to make many changes. Due to the coupling between the transmitter and the receiver, throughput is most likely to achieve an average expected triple.

At first, the receiver needs to estimate and correct the frequency offset and symbol timing using the training symbols in the preamble coding. Subsequently, the CP is removed and Discrete Fourier transform (DFT) is performed for each receiver link. It is necessary to perform MIMO detection for each OFDM subcarrier. So that the received signal of the each subcarrier is routed to the MIMO detector, thereby recovering the data signal transmitted on the subcarrier. Next, the symbols of each data stream are combined, later the parallel streams are de-mapped, sorted, and decoded, and the resultant data is finally combined to obtain binary output data [10].

Figure 7 is a system architecture diagram of MIMO-OFDM. The following figure shows the symbol error rate and SNR of MIMO-OFDM for arbitrary transmit and receive antennas. As can be seen from Fig. 8, in the MIMO-OFDM system, with the signal-to-noise ratio change, the bit error rate of the channel is gradually reduced.

5 Conclusion

OFDM is a high-efficient modulation scheme, and it is widely used in various fields of broadband digital communication because of its superior anti-multipath capability and spectrum utilization rate. MIMO technology can obtain the diversity gain through the joint coding between the multi-pair transmitting antennas, thus improving the reliability, and can obtain the doubling channel capacity by spatial multiplexing. MIMO technology's multipath fading robustness combined with OFDM can provide high data rates and spectral efficiency, as well as enhance link reliability, improve energy efficiency.

Acknowledgments. This work was supported by Hainan Provincial Key R. & D. Projects of China (ZDYF2016010 and ZDYF2018012) and the National Natural Science Foundation of China (No. 61661018).

References

1. Michailidis, E.T., Theofilakos, P., Kanatas, A.G.: Three-dimensional modeling and simulation of mimo mobile-to-mobile via stratospheric relay fading channels. IEEE Trans. Veh. Technol. **62**(5), 2014–2030 (2013)
2. Pätzold, M., Hogstad, B.O., Youssef, N.: Modeling, analysis, and simulation of MIMO mobile-to-mobile fading channels. IEEE Trans. Wirel. Commun. **7**(2), 510–520 (2008)
3. Yong, S.C., Kim, J, Yang, W.Y., Kang, C.G.: MIMO-OFDM Wireless Communications with MATLAB. Wiley Publishing (2010)
4. Guan, H.: Key technology to 5G network. In: Asia-Pacific Microwave Conference 2015, pp. 1 (2015)

5. Koffman, I., Roman, V.: Broadband wireless access solutions based on OFDM access in IEEE 802.16. IEEE Commun. Mag. **40**(4), 96–103 (2013)
6. Zhang, W., Li, H., Li, B.: Iterative decision-directed channel estimation for MIMO-OFDM system. In: 2016 2nd IEEE International Conference on Computer and Communications, pp. 1678–1682 (2016)
7. Kowal, M., Kubal, S., Piotrowski, P., Zielinski R.J.: Simulation model of the MIMO-OFDM system compliant with IEEE 802.11n. In: 2010 5th International Conference on Broadband and Biomedical Communications, pp. 1–5 (2010)
8. Tarokh, V., Jafarkhani, H., Calderbank, A.R.: Space-time block codes from orthogonal designs. IEEE Trans. Inf. Theory **45**(5), 1456–1467 (1999)
9. Sandhu, S., Paulraj, A.: Space-time block codes: a capacity perspective. IEEE Commun. Lett. **4**(12), 384–386 (2000)
10. Hsu, C.Y., Yeoh, P.L., Krongold, B.S.: Power minimization in cooperative MIMO-OFDM systems with user fairness constraints. In: 2015 IEEE Wireless Communications and Networking Conference, WCNC 2015, pp. 93–98 (2015)

Subcarrier Allocation of Cognitive OFDM Systems with SWIPT-Enabled Relay

Yuanrong Lin[1], Weidang Lu[1(✉)], Qibin Ye[1], Hong Peng[1], Yu Zhan[1], and Xin Liu[2]

[1] College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China
luweid@zjut.edu.cn
[2] Dalian University of Technology, Dalian 116024, China

Abstract. Cognitive radio (CR), as an intelligent spectrum sharing technology, can improve utilization of spectrum. Simultaneous wireless information and power transfer (SWIPT) is harvesting energy from ambient RF signals and makes full use of RF signals. Orthogonal frequency division multiplexing (OFDM) can allocate subcarrier flexibly and the transmission power on the subcarrier can be controlled. In this paper, we utilize subcarrier allocation method to dispose the interface at receiving end for cognitive OFDM systems.

Keywords: Simultaneous wireless information and power transfer · Subcarrier allocation · Decode-and-forward · Orthogonal frequency division multiplexing

1 Introduction

In recent years, spectrum scarcity becomes the bottleneck for the development of wireless communication [1]. Cognitive radio (CR), as an intelligent spectrum sharing technology, can improve the utilization of spectrum [2, 3]. Radio frequency (RF) signal, is an emerging resource which can carry energy and information at the same time. Simultaneous wireless information and power transfer (SWIPT) is harvesting energy from ambient RF signals [4]. Orthogonal frequency division multiplexing (OFDM) can allocate subcarrier flexibly and the transmission power on the subcarrier can be controlled [5]. In this paper, we utilize subcarrier allocation method to dispose the interface at receiving end for cognitive OFDM systems.

2 System Model

As illustrated in Fig. 1, we consider CRNs where a primary network coexists with a secondary network. There are a pair of PUs which are, respectively, primary transmitter (PT) and primary receiver (PR). The secondary network consists of three nodes, i.e., a source node (SN), a relay node (RN), and a destination node (DN), respectively. RN utilizes power splitting (PS) architecture with power splitting ratio $\lambda(0 \leq \lambda \leq 1)$. We

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 317–320, 2019.
https://doi.org/10.1007/978-981-13-6264-4_38

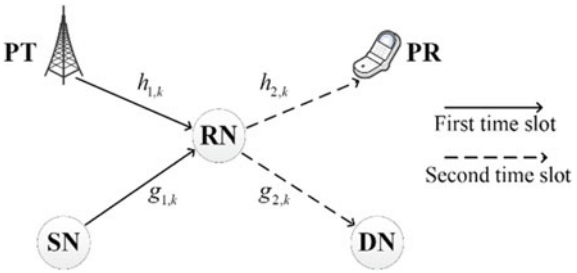


Fig. 1. System model.

denote the channel gains of the k^{th} subcarrier over links $PT \rightarrow RN$, $RN \rightarrow PR$, $SN \rightarrow RN$, and $RN \rightarrow DN$ as $h_{1,k}$, $h_{2,k}$, $g_{1,k}$, and $g_{2,k}$.

In the first phase, PT transmits the signal $x_{p,k}$ to RN while SN transmits the signal $x_{s,k}$ to RN over subcarrier k, the transmission power over subcarrier k of PT $p_{p,k}$ and SN $p_{s,k}$ is obtained by water-filling algorithm, The baseband signal of the k^{th} subcarrier at the information receiver is

$$y_k^{RN,R} = \sqrt{\lambda}\left(\sqrt{p_{p,k}}h_{1,k}x_{p,k} + \sqrt{p_{s,k}}g_{1,k}x_{s,k} + n_{1,k}\right) + n_{2,k}, \tag{2}$$

where $n_{1,k} \sim N(0, \sigma_{1,k}^2)$ is the additive white Gaussian noise (AWGN) at RN, $n_{2,k} \sim N(0, \sigma_{2,k}^2)$ is the noise generated in information receiver [6]. Thus, the throughput over link $SN \rightarrow RN$ in the second phase is expressed as

$$R_{SN-RN} = \sum_{k \in K} \frac{1}{2} \log_2 \left(1 + \frac{\lambda p_{s,k} \left|g_{1,k}\right|^2}{\lambda p_{p,k} \left|h_{1,k}\right|^2 + \lambda \sigma_{1,k}^2 + \sigma_{2,k}^2}\right). \tag{3}$$

In the second phase, RN uses a part of subcarriers for relaying PT's information to PR and the rest of subcarriers G for relaying SN's information to DN. The received signal at DN is

$$y^{DN} = \sum_{k' \in G} \left(\sqrt{p_{rs,k'}} g_{2,k'} x_{s,k'} + n_{3,k'}\right), \tag{5}$$

where $n_{3,k} \sim N(0, \sigma_{3,k}^2)$ is the AWGN at DN. Thus, the throughput over link $RN \rightarrow DN$ in the second phase is expressed as

$$R_{RN-DN} = \sum_{k' \in G} \frac{1}{2} \log_2 \left(1 + \frac{p_{rs,k'} \left|g_{2,k'}\right|^2}{\sigma_{3,k'}^2}\right). \tag{10}$$

Since the relaying protocol we utilized is DF, the throughput of secondary network is

$$R_{SU} = \min(R_{SN-RN}, R_{RN-DN}).$$

3 Problem Formulation and Solution

We aim at maximizing R_{SU} while guaranteeing the communication performance of primary network. For simplicity, the transmission power of each subcarrier at RN utilizes equal division. First, RN distributes subcarriers for relaying PT's information to PR until the throughput over link $RN \rightarrow PR$ achieves target RT. Second, RN utilizes the rest of subcarriers for relaying SN's information to DN.

4 Simulation Results and Conclusion

We assume that the path loss exponent $m = 3$, the distance $d_{SN,RN} = d_{RN,DN} = 1.5$, $d_{\mathrm{PT},RN} = d_{\mathrm{RN},PR} = 2$, the energy harvesting efficiency $\eta = 0.8$, and the total power of PT is set to $P_{\mathrm{PT}} = 2\mathrm{W}$. For simplicity, the power of noise is set to $\sigma_a^2 = \sigma_b^2 = \sigma_c^2 = 0.0001$. Figure 2 shows the throughput of secondary network versus the total power of SN P_{SN} with different RT. λ is set to be 0.5. In Fig. 2, we can find that with lower RT, more power can be used for transmitting SN's information, so that secondary network has greater throughput. Figure 3 presents the throughput versus the P_{SN} with different λ. And, $RT = 1\,\mathrm{bps}/HZ$. Figure 3 shows that the throughput for λ^* which is obtained by simulation is greater than other λ.

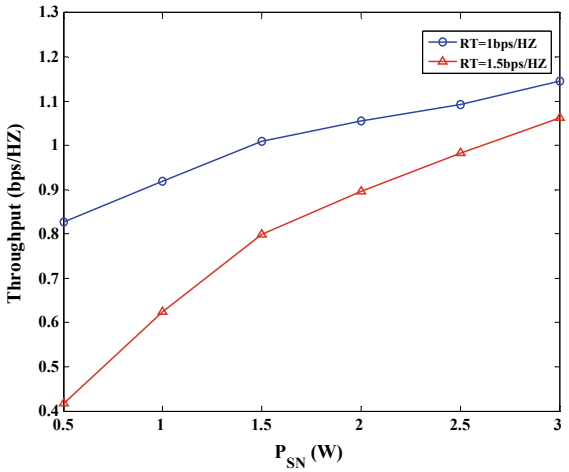


Fig. 2. Throughput versus P_{SN}.

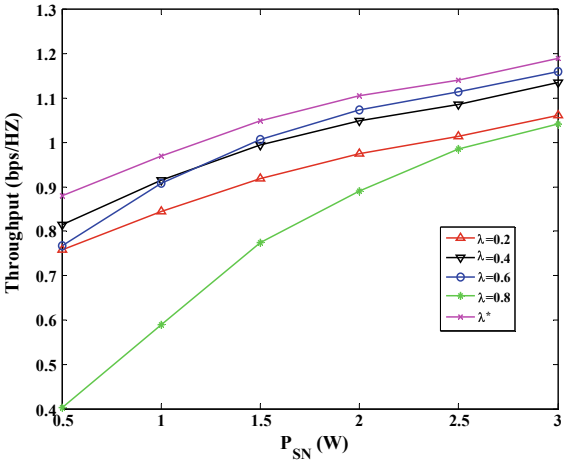


Fig. 3. Throughput for different λ versus P_{SN}.

In this paper, we utilize subcarrier allocation method to dispose the interface at receiving end for cognitive OFDM systems and the simulation results present the performance of our method.

References

1. Wang, J., Ghosh, M., Challapali, K.: Emerging cognitive radio applications: a survey. IEEE Commun. Mag. **49**(3), 74–81 (2011)
2. Haykin, S.: Cognitive radio: brain-empowered wireless communications. IEEE J. Sel. Areas Commun. **23**(2), 201–220 (2005)
3. Wang, B., Liu, K.: Advances in cognitive radio networks: a survey. IEEE J. Sel. Top. Signal Process. **5**(1), 5–23 (2011)
4. Lu, X., Wang, P., Niyato, D., Kim, D.I., Han, Z.: Wireless networks with RF energy harvesting: a contemporary survey. IEEE Commun. Surv. Tuts. **17**(2), 757–789 (2015)
5. Wu, Y., Zou, W.Y.: Orthogonal frequency division multiplexing: a multi-carrier modulation scheme. IEEE Trans. Consum. Electron. **41**(3), 392–399 (1995)
6. Nasir, A.A., Zhou, X., Durrani, S., Kennedy, R.A.: Relaying protocols for wireless energy harvesting and information processing. IEEE Trans. Wireless Commun. **12**(7), 3622–3636 (2013)

Resource Allocation for Energy Efficiency Maximization in OFDM-Based SWIPT Systems

Shanzhen Fang[1], Weidang Lu[1(✉)], Su Hu[2], Xin Liu[3], Zhenyu Na[4], and Jingyu Hua[1]

[1] College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China
luweid@zjut.edu.cn
[2] University of Electronic Science and Technology of China, Chengdu 611731, China
[3] Dalian University of Technology, Dalian 116024, China
[4] Dalian Maritime University, Dalian 116026, China

Abstract. In this paper, we propose a resource allocation algorithm for maximizing system energy efficiency in a downlink OFDM-based SWIPT system, in which the energy efficiency of the system is maximized by jointly optimizing the subcarrier set and power allocation. The original optimization problem is difficult to solve directly. By transforming the objective function, we use the Lagrangian duality method to obtain the optimal solution. Simulation results show that our proposed algorithm is superior to the current algorithms.

Keywords: Energy efficiency · Power and subcarrier allocation · SWIPT · OFDM

1 Introduction

Traditional wireless communication research is mainly focused on maximizing the achieved rate of the system [1]. However, while achieving a large rate, it may consume a lot of power, resulting in low energy efficiency. So more and more scholars are studying the maximum energy efficiency of the system while achieving the target rate of the system. Reference [2] proposed a dynamic power splitting (DPS) scheme to improve the system energy efficiency in clustered wireless sensor networks. Reference [3] studied two cases, when the transmission power is a discrete set of power and a continuous set of power, respectively, the energy efficiency is optimized in these two cases. Reference [4, 5] applied the widely used Dinkelbach iterative algorithm to improve the energy efficiency.

In this paper, we propose a new resource allocation algorithm that maximizes energy efficiency while achieving the target rate and minimum collected energy.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 321–325, 2019.
https://doi.org/10.1007/978-981-13-6264-4_39

2 System Model and Problem Formulation

2.1 System Model

In a downlink OFDM single-user system, there are N subcarriers. S^P represents the set of subcarriers used to collect energy and S^I represents the set of subcarriers used to decode the information. The power of the user on subcarrier n is expressed as p_n. We specify that the user has a minimum target rate R, and B denotes the minimum required energy for the user. The total power of the system is set to P, and the channel gain is denoted as h_n, as shown in Fig. 1.

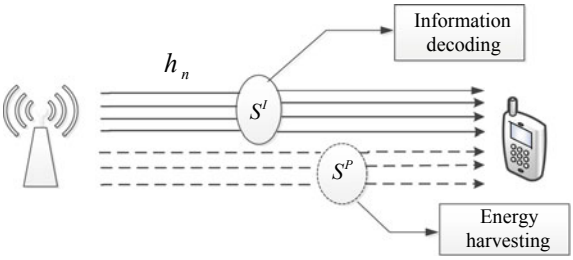


Fig. 1. System model

Thus, the achievable rate on subcarrier n is $\log(1+\frac{h_n p_n}{\sigma^2})$. Thus, the weighted system throughput is given as

$$U(P,S)=\sum_{n\in S^I}\log(1+\frac{h_n p_n}{\sigma^2}). \tag{1}$$

Similarly, the energy collected by the user on subcarrier n is $\varepsilon h_n p_n+\sigma^2$, the weighted power consumption is expressed as

$$U_{TP}(P,S)=P_B+P_R+\sum_{n\in N}p_n-\sum_{n\in S^P}(\varepsilon h_n p_n+\sigma^2), \tag{2}$$

where P_B and P_R represent the fixed power consumption of the transmitter and receiver, respectively, and ε denotes energy harvesting efficiency. According to [3, 4], we can easily get the expression of system energy efficiency as

$$U_{eff}(P,S)=\frac{U(P,S)}{U_{TP}(P,S)}. \tag{3}$$

2.2 Problem Formulation

The optimization problem can be formulated as

$$\begin{aligned}
&\max_{p_n, S^P} U_{eff}(P, S) \\
&s.t.\ C1: \sum_{n \in S^I} \log(1 + \frac{h_n p_n}{\sigma^2}) \geq R,\ \text{C2:} \sum_{n \in S^P} (\varepsilon h_n p_n + \sigma^2) \geq B \\
&C3: P_B + P_R + \sum_{n \in N} p_n \leq P,\ \text{C4:}\ S^I + S^P = N \\
&C5: S^I \cap S^P = \varnothing,\ \text{C6:}\ p_n \geq 0.
\end{aligned} \tag{4}$$

3 Optimal Solution

The optimal resource allocation is obtained through the following two steps; the first step is obtaining the optimal power for fixed subcarrier set.

Since the objective function is fractional, we convert the objective function into $U(P, S) - qU_{TP}(P, S)$. The Lagrange equation of the transformed optimization problem can be expressed as

$$\begin{aligned}
L(p, \beta) = &\sum_{n \in S^I} \log(1 + \frac{h_n p_n}{\sigma^2}) - q[P_B + P_R + \sum_{n \in N} p_n - \sum_{n \in S^P} (\varepsilon h_n p_n + \sigma^2)] \\
&+ \beta_1 [\sum_{n \in S^I} \log(1 + \frac{h_n p_n}{\sigma^2}) - R] + \beta_2 [\sum_{n \in S^P} (\varepsilon h_n p_n + \sigma^2) - B] \\
&+ \beta_3 [P - \sum_{n \in N} p_n - P_B - P_R],
\end{aligned} \tag{5}$$

where $\{\beta_1, \beta_2, \beta_3\}$ are nonnegative dual multipliers, and can be solved using subgradient method [6]. After a series of complicated calculations, the optimal power can be obtained as

$$p_n^*(n \in S^I) = [\frac{(1 + \beta_1)}{q + \beta_3} - \frac{\sigma^2}{h_n}]^+ \tag{6}$$

$$p_n^*(n \in S^P) = \begin{cases} p_{\max}, (q\varepsilon h_n + \varepsilon \beta_2 h_n) > \beta_3 + q \\ p_{\min}, otherwise \end{cases}, \tag{7}$$

where $p_{\max}$ and $p_{\min}$ represent the peak and lowest power constraints, and $[x]^+ \triangleq \max\{0, x\}$.

Next, let us find the optimal subcarrier set. Substituting (6) and (7) into (5), the original equation can be transformed into

$$L(p,\beta) = \sum_{n\in N} log(1+\frac{h_n p_n^*}{\sigma^2}) + \sum_{n\in N} \beta_1 \log(1+\frac{h_n p_n^*}{\sigma^2}) + \sum_{n\in S^P} F$$
$$- \beta_3 \sum_{n\in N} p_n^* - q \sum_{n\in N} p_n^* - q(P_B + P_R) - \beta_1 R - \beta_2 B + \beta_3 P - \beta_3 (P_B + P_R), \tag{8}$$

where $F = \beta_2(\varepsilon h_n p_n^* + \sigma^2) + q(\varepsilon h_n p_n^* + \sigma^2) - \beta_1 \log(1+\frac{h_n p_n^*}{\sigma^2}) - \log(1+\frac{h_n p_n^*}{\sigma^2})$.

$$S^{P*} = \arg\max \sum_{n\in S^P} F. \tag{9}$$

Thus, the remaining subcarriers belong to S^{I*}.

4 Simulation Results and Conclusions

The figure on the left shows that as the total power increases, the energy efficiency of the system also increases. When the user reaches the target rate and the minimum collected energy, the total power increases, and the energy efficiency will no longer increase. In Algorithm 1, the power is divided equally over each subcarrier and in Algorithm 2, subcarriers sets are fixed, and the number of subcarriers set for information decoding and energy collection is the same. We can observe from the figure on the right that with the increase of P_B, the energy efficiency of the system gradually decreases. It is clear that our proposed algorithm is superior to these two algorithms (Fig. 2).

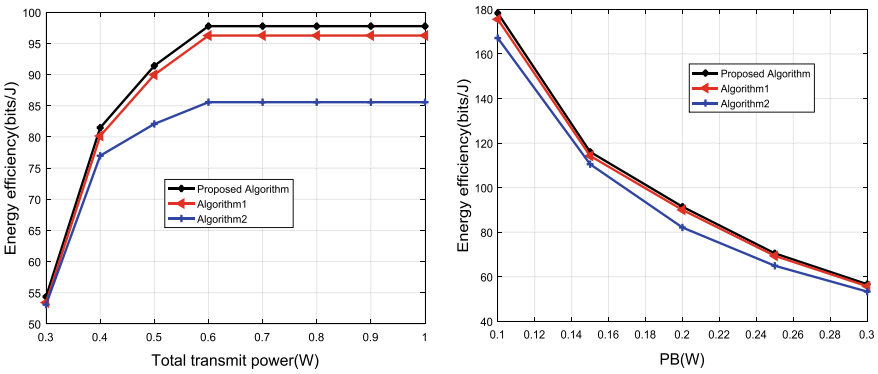


Fig. 2. Energy efficiency versus total transmit power and P_B

References

1. Lu, W.D., Gong, Y., Liu, X., Wu, J.Y., Peng, H.: Collaborative energy and information transfer in green wireless sensor networks for smart cities. IEEE Trans. Ind. Inf. **14**(4), 1585–1593 (2018)
2. Guo, S., Wang, F., Yang, Y., Xiao, B.: Energy-efficient cooperative transmission for simultaneous wireless information and power transfer in clustered wireless sensor networks. IEEE Trans. Commun. **63**(11), 4405–4417 (2015)
3. Ng, D.W.K., Lo, E.S., Schober, R.: Wireless information and power transfer: energy efficiency optimization in OFDMA systems. IEEE Trans. Wirel. Commun. **57**(6), 6352–6370 (2013)
4. Wu, Q., Chen, W., Tao, M., Li, J., Tang, H., Wu, J.: Resource allocation for joint transmitter and receiver energy efficiency maximization in downlink OFDMA systems, IEEE Trans. Commun. **63**(2), 416–430 (2015)
5. Ng, D.W.K., Lo, E.S., Schober, R.: Energy-efficient resource allocation in OFDMA systems with large numbers of base station antennas. IEEE Trans. Wirel. Commun. **11**(9), 3292–3304 (2012)
6. Boyd, S., Mutapcic, A.: Sub-gradient methods, Notes for EE364. Standford University, Winter 2006–07

Energy Trading Scheme Based on Stackelberg Game to Optimize Energy Price and Power

Chen Liu, Weidang Lu(✉), Jie Zhang, Hong Peng, Zhijiang Xu, and Jingyu Hua

College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China
luweid@zjut.edu.cn

Abstract. In this article, we use the Stackelberg game based on energy trading scheme to optimize two variable parameters at the same time. In wireless RF energy harvesting, we use energy access point EAP to charge the relay to help the system transfer information. Finally, the destination node will pay corresponding rewards to the EAP.

Keywords: Stackelberg game · Energy harvesting · Energy trading · Optimize parameters

1 Introduction

In recent years, wireless charging technology has become a hot trend in communication. [1–3] introduces incentives mechanism for spectrum bandwidth sharing and energy trading, respectively. Energy collection and resource allocation in wireless networks are introduced in [4, 5].

This paper introduces the Stackelberg game for energy trading with external EAP. Finally, the simulation results show the relationship between the optimal energy price and other parameters.

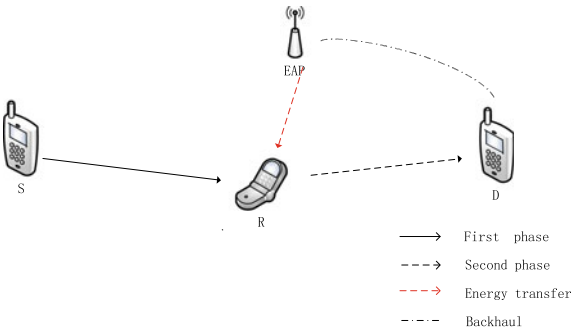


Fig. 1. System model

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 326–330, 2019.
https://doi.org/10.1007/978-981-13-6264-4_40

2 System Model and Problem Formulation

As shown in Fig. 1, we suppose S to forward information to D through a relay R by AF protocol, but R's energy is empty. Surrounding EAP provides energy to R to help R to forward information. Then, EAP acquires the profit by the backhaul.

The energy received by R from the EAP is expressed as

$$E_R = \eta p_E h_{E,R} \tag{1}$$

where p_E is the charging power of the EAP and $h_{E,R}$ is the channel power gain between the EAP and the relay node.

We assume that the R uses all the energy to forward the information. Therefore, the energy received by the relay from the EAP is equal to the transmission power of the relay forwarding information to D, i.e., (P_R).

The system uses AF protocol to forward message.

$$y_1 = \sqrt{P_S} h_{S,R} x_1 + n_1 \tag{2}$$

$$y_2 = \Phi y_1 h_{R,D} + n_2 \tag{3}$$

Therefore, the SNR of destination node is expressed as

$$SNR = \frac{\frac{p_S |h_{S,R}|^2}{\sigma^2} \cdot \frac{\eta p_E h_{E,R} |h_{R,D}|^2}{\sigma^2}}{1 + \frac{p_S |h_{S,R}|^2}{\sigma^2} + \frac{\eta p_E h_{E,R} |h_{R,D}|^2}{\sigma^2}} \tag{4}$$

The achievable rate at the D can be expressed as

$$R_D = \frac{1}{2} \log_2 (1 + \frac{\eta p_S |h_{S,R}|^2 |h_{R,D}|^2 p_E h_{E,R}}{\sigma^4 + \sigma^2 p_S |h_{S,R}|^2 + \eta \sigma^2 |h_{R,D}|^2 p_E h_{E,R}}) \tag{5}$$

Now, make a replacement

$$q_R = p_E h_{E,R} \tag{6}$$

Therefore,

$$R_D = \frac{1}{2} log_2 \left(1 + \frac{\mathrm{a} q_R}{\mathrm{b} + \mathrm{c} q_R} \right) \tag{7}$$

The reward for EAP is

$$\Lambda(\lambda, q) = \lambda_E q_R \tag{8}$$

λ_E denoting the price per unit energy harvested from the EAP.

$$C_E(x) = a_E x^2 = \frac{a_E}{h_{E,R}^2} q_R^2 = \frac{q_R^2}{\theta} \tag{9}$$

$C_E(x)$ indicates the energy cost of the EAP.
The utility function of the D and the EAP can be, respectively, defined as

$$\begin{cases} U_D(\lambda, q) = R_D - \lambda_E q_R \\ U_E(\lambda, q) = \lambda_E q_R - \dfrac{q_R^2}{\theta} \end{cases} \tag{10}$$

So, the optimization problem can be formulated as

$$\begin{aligned} &\mathbf{P}.\mathbf{1} : \max\{U_D(\lambda_E, q_R)\} \text{ s.t } \lambda_E \geq 0 \\ &\mathbf{P}.\mathbf{2} : \max\{U_E(\lambda_E, q_R)\} \text{ s.t } q_E \geq 0 \end{aligned} \tag{11}$$

3 Optimal Solution

Lemma 1 For given values of λ_E, we can get the optimal value of q_R.

$$q_R^* = \frac{\lambda_E \theta}{2} \tag{12}$$

Proof The objective function is a quadratic function about q_R.

Bring q_R^* into Eq. (11), Then, we can get the optimal solution of λ_E when the first derivation of U_D equals to zero. Simplified as follows:

$$\lambda_E^3 (\frac{1}{4} c^2 \theta^2 + \frac{1}{4} ac\theta^2) + \lambda_E^2 (bc\theta + \frac{1}{2} ab\theta) + \lambda_E b^2 - \frac{ab}{4 \ln 2} = 0 \tag{13}$$

where

$$a' = \frac{1}{4} c^2 \theta^2 + \frac{1}{4} ac\theta^2, b' = bc\theta + \frac{1}{2} ab\theta, c' = b^2, d' = -\frac{ab}{4 \ln 2} \tag{14}$$

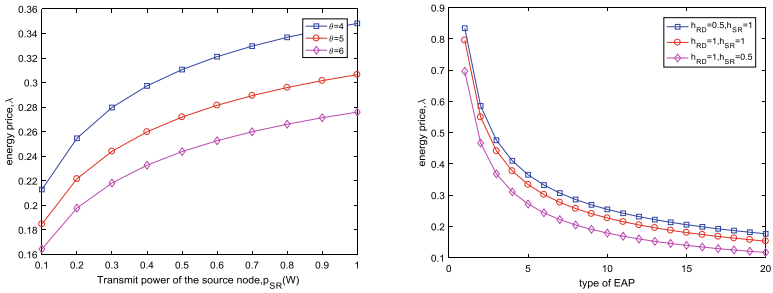
$$A = b'^2 - 3a'c', B = b'c' - 9a'd', C = c'^2 - 3b'd' \tag{15}$$

The cubic equation discriminant is $\Delta = B^2 - 4AC > 0$

$$\begin{cases} X_1 = \dfrac{-b' - (\sqrt[3]{Y_1} + \sqrt[3]{Y_2})}{3a'} \\ X_{2,3} = \dfrac{-b' + \frac{1}{2}(\sqrt[3]{Y_1} + \sqrt[3]{Y_2}) \pm \frac{\sqrt{3}}{2}(\sqrt[3]{Y_1} - \sqrt[3]{Y_2})i}{3a'} \end{cases} \tag{16}$$

$$Y_{1,2} = Ab' + 3a'\left(\frac{-B \pm \sqrt{B^2 - 4AC}}{2}\right) \tag{17}$$

4 Simulation Results



According to the above two figures, the left shows that the optimal energy price will increase as the transmit power of S increases. This is because the transmit power of S becomes larger and requires more energy for transmitting information.

From the right figures, we can see that the higher the type of EAP, the lower the energy price. This is because the higher the type, the better the channel quality will be. We can also see energy price is related to channel gain per time slot.

5 Conclusion

In this article, we introduced a Stackelberg game about radio frequency energy harvesting. We also optimized two parameters, EAP's energy price and D's received power, so that they can reach their optimal values in their respective systems. In the game theory of this paper, our goal is to find a balance point that maximizes their profits. The simulation numerical results show the relationship between the optimal value and other constants.

References

1. Hou, Z., Chen, H., Li, Y., Vucetic, B.: Incentive mechanism design for wireless energy harvesting-based internet of things. IEEE Internet Things J. (2017). doi: https://doi.org/10.1109/JIOT
2. Lu, W., He, C., Lin, Y.: Contract theory based cooperative spectrum sharing with joint power and bandwidth optimization, project funded by China Postdoctoral Science Foundation under Grand No. 2017M612027
3. Chen, H., Li, Y., Han, Z., Vucetic, B.: A stackelberg game-based energy trading scheme for power beacon-assisted wireless-powered communication. In: Proc. ICASSP, pp. 3177–3181 (2015)
4. Xu, Y., Mao, S.: Stackelberg game for cognitive radio networks with mimo and distributed interference alignment. IEEE Trans. Veh. Technol. **63**(2), 879–892 (2014)
5. Lu, X., Wang, P., Niyato, D., Han, Z.: Resource allocation in wireless networks with RF energy harvesting and transfer. IEEE Netw. (2014)

Wireless Networks

HCR-MAC: A High-Capacity Receiver-Based MAC Protocol for Cognitive Radio Enabled AMI Networks in Smart Grid

Zhuoran Cai(✉), Zhichao Qu, and Xiao Zhang

Shandong Institute of Space Electronic Technology, Yantai 264670, China
qingdaogancai@126.com, moyuertalang@163.com, x.zhang.cn@ieee.org

Abstract. For the purpose of enhancing electric power grid in a variety of application scenarios and realizing smart grid, the Cognitive Radio Enabled Advanced Metering Infrastructure (CR-AMI) network is recognized as an effective solution. However, there have been many evitable challenges in advanced smart grid as a result of the harsh wireless environment. Therefore, two crucial issues should be focused on are throughput and reliability. In this regard, we propose a high-capacity receiver-based MAC protocol for cognitive AMI, which is termed as HCR-MAC. In order to achieve high capacity and reliability, HCR-MAC adopts a capacity-based auction mechanism and preamble sampling techniques respectively. The proposed protocol also clearly accounts for the properties of a CR environment. Simulation results demonstrate HCR-MAC is able to provide an effective solution for cognitive MAI networks in smart grid.

Keywords: Smart grid · Advanced Metering Infrastructure · Cognitive radio · MAC protocol · Capacity

1 Introduction

Some intrinsic defects in the legacy electric power grid, such as insecurity, energy inefficiency, and congestion in transmission [1], has not been capable to meet increasing demands well. Smart grid, which is proposed to replace the current electric grid, and serve as the next generation, has the potential to provide reliable, energy-efficient, agile and secure services [2,3]. Smart grid is expected to have some capabilities, such as advanced bi-directional communications, automatic control, distributed computing, and so on. Besides, smart grid can gather information from equipment in different locations in real-time, followed by adopting reasonable strategies to achieve expected energy efficiency and reliability. Therefore, it will be beneficial for all participants in electric grid to get along with real-time cognitive services, such as operating environments, requirement response, and so on.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 333–341, 2019.
https://doi.org/10.1007/978-981-13-6264-4_41

Advanced Metering Infrastructure (AMI) is a key element for information gathering and processing in the next generation electric grid. In AMI networks, smart meters equipped by each consumer are abided by the Meter Data Management System (MDMS) together, which serves as the control center to store and manage power-consuming information of consumers. These information can be used by different applications [4]. Furthermore, the AMI network, which can perform mutual communications between utilities and customers, are dispensable in smart grids, serving electricity producers, suppliers, and users. The suppliers can supervise the distribution of electricity usage and service quality for consumers. The consumers can explore the real-time electricity price and on this basis, consume power opportunistically. However, there is a bottleneck to traditional AMI networks result from the inefficiency and scarcity of spectrum resources. Cognitive radio (CR) can serve as an effective solution for the inefficiency and scarcity issues in AMI applications. Literature [5] has shown a practical method to address the mentioned spectrum resource issues in wireless network, which is also important for smart grid generalization. Therefore, the combination of cognitive radio and AMI networks becomes the research hotspot. Researchers expect the smart meters can use the channel and frequency resource when the primary users are absence, or smart meters can occupy the channel without interferences to the communications of licensed users. Loads of studies have proposed a variety of CR-AMI networks involving CR techniques into smart grid, such as [3,6].

A proposed cognitive radio enabled wireless mesh network can connect numerous smart meters to MDMS, which provides a practical way to enhance AMI networks. Smart meters transmit their meter data to the gateway in turn [7]. However, in harsh environmental conditions, the successful transmission in smart grid is determined by the intrinsic communication capabilities of AMI networks. To improve the energy and spectrum efficiency of data transmission in smart grids, improvements should be imposed on different layers of protocol stacks, including the Medium Access Control (MAC) layer. In [6], authors show a receiver-based MAC protocol. It improves the data transmission reliability, and gets a trade-off between the energy efficiency and spectrum utilization. Nevertheless, the throughput of communication is not optimized in this work.

Thereby, we focus on improving the throughput capability during the enhancement of receiver-based MAC protocol in the CR-AMI network. In this regard, the protocol is proposed and termed as HCR-MAC, which is *receiver-based* in nature. The objective of HCR-MAC is improving throughput of CR-AMI network in the next generation grid environments. In HCR-MAC, a high throughput can be obtained by refer to the capacity as the crucial factor in the next-hop competition. Furthermore, HCR-MAC deploys preamble sampling [3] method to address with idle listening and sleep/wakeup modes with no requirements for synchronization overheads. Especially, HCR-MAC adopts a multi-receivers auction mechanism by exploiting the broadcast property of wireless transmission in the later sections.

The remainder of this paper is arranged as follows. We will introduce the framework of HCR-MAC and the system model in Sect. 2. In Sect. 3, simulation performance evaluation and numerical analytical are shown. Finally, Sect. 4 will conclude this work.

2 HCR-MAC Framework

2.1 HCR-MAC Overview

In HCR-MAC, there exist multiple receiver nodes for a single hop transmission. Hence, it belongs to receiver-based MAC protocols intrinsically. Due to the receiver-based nature, the protocol is distinct from *sender-based* MAC protocol, where the sender defines the receiver node. It selects the receiver from its forwarder list based on the specific rules, and adds the receiver name into the preamble of data packets. While for HCR-MAC protocol, there is no one specific node selected to serve as the appointed receiver. If any node wants to transmit data packets, it can carry out in a broadcasting way. All of the near nodes locating within the communication coverage will find the data packet. Receivers participate in competition under an election mechanism, where the final winner will forward the packet by broadcasting.

A sender node broadcasts its packet instead of selecting a specific node as the receiver, and all neighbor nodes of the sender node are able to receive the data packet. Each individual receiver checks its eligibility to forward the data according to the information of the received preamble. Receivers participate in the elective competition and the winner will forward the packet to the next hop toward gateway/sink. Note that the capacity is the key factor of the elective process. During the process, any node contending to forward data estimates the capacity of single hop operation. Based on the estimated value of capacity, the duration before forwarding operation is determined. As a result, the receiver node with better capacity during this hop will have higher chance to forward the data packet.

Another innovative aspect of HCR-MAC protocol is employing preamble sampling to enhance the energy efficiency. In this approach, nodes exploit listen with lower power asynchronously, and determine the sleep/wakeup schedules independently. Most nodes stay in sleep mode at most of their time and only work for a short time. To minimize the probability of missed detection, we adopt the preamble, which is inserted ahead of the data packet. The preamble has a greater length than checking interval (CI) to ensure itself can be inspected. Owing to adjusting CI and CCA, average duty-cycles can be reduced below 1% without presence of scheduling and synchronization.

As an essential section of CR technologies, all smart meters in HCR-MAC is assumed to have the ability of spectrum sensing. Nodes detect the channel state and primary user (PU) activities periodically, and especially before using it for transmission of cognitive users. In sensing process, nodes cannot forward data packets leading to degradation of the network performance (e.g., in accordance with throughput and end-to-end delay). HCR-MAC adopts a mechanism

to enhance the overall network performance in terms of spectrum sensing processes of each node. In addition, HCR-MAC utilizes the optimized transmission duration subject to an interference constraint to avoid influencing PUs.

2.2 System Model

The multi-hop mesh network with stationary CR enabled smart meters is considered in this work. We assume that there are J PUs, which are stationary in a region. The locations and communication ranges of PU transmitters are known. The activity of each PU is modeled by using a two-state independent and identically distributed random process. We assume the two states are *idle* and *busy*, and the states of channel j are described by S_b^j and S_i^j, where S_b^j denotes the busy state of channel j, and S_i^j denotes the idle state of channel j. We suppose that each smart meter can utilizes energy detection method for primary user detection, by comparing the received signal energy (E) with the predetermined threshold (σ), in order to determine whether the j^{th} channel is available, expressed as follows.

$$Result_{SS} = \begin{cases} S_{busy}^i \text{ if } & E \geq \sigma \\ S_{idle}^i \text{ if } & E < \sigma \end{cases} \tag{1}$$

The structure of a MAC frame in the AMI network is composed of a spectrum sensing slot (T_s) and a transmission slot (T). There are two primary metrics in spectrum sensing, P_d and P_f, which denote the probability of detection and the probability of false alarm, respectively. The former ensures improved protection to occupants, while the later guarantees better usage of the channel. Both detection and false alarm probabilities are given by

$$P_f^j = Pr\{E \geq \sigma | S_i^j\} = Q\left(\frac{\sigma - 2n_j}{\sqrt{4n_j}}\right), \tag{2}$$

$$P_d^j = Pr\{E \geq \sigma | S_i^j\} = Q\left(\frac{\sigma - 2n_j\left(\gamma_j + 1\right)}{\sqrt{4n_j\left(2\gamma_j + 1\right)}}\right), \tag{3}$$

where $Q(\cdot)$ denotes the Q function, and γ_j denotes the SNR of the primary user signal, and n_j denotes the bandwidth-time product for the j^{th} channel.

2.3 Protocol Description

In protocol HCR-MAC, compared with the *sender-based* mechanism (for example, the protocol in [7]), it is unnecessary for sender nodes to determine a specific receiver. Instead, the source node of a transmission will transmit data packets by broadcasting, and the neighbor nodes, which receive the packets, determine the forwarder node. The nodes receiving packets determine the forwarder node. The HCR-MAC protocol single hop operation and the schedule for different nodes are shown as Fig. 1. Node S has data to send to the gateway. Firstly, spectrum

sensing is carried out in the duration T_s, in order to find whether there exists any PU in channels.When the PU is not detected, namely, S_i^j, S gets down to broadcast the preamble. The preamble with duration of T_{pr}, consists of multiple micro-frames, the identification information of which can help neighboring nodes to classify transmission between PU and cognitive nodes. The PU node broadcast a group of micro-frames as the preamble to all of its neighboring nodes in the range of S. Suppose that nodes A, B, C are three neighboring nodes in coverage of S and qualified to deliver the data packet toward the gateway. When they listen to the preamble from S, A, B, C wake up and begin to receive the packet. If there is any error in the data in the packet, nodes A, B, C will give up the forwarding. For nodes intending to forward the packet, they need wait for a timer Δt_X, where X represents the label of the forwarder node. After the duration, they can carry out the spectrum sensing. When there is a idle channel, is broadcasts the preamble for forwarding. The first node which broadcasts the preamble is the winner and will forward the packet. The duration Δt of node y is given by

$$\Delta t_y = \omega_0 + (\omega_1 C_{x,y}^j)^{-1}, \tag{4}$$

where ω_0 and ω_1 are constants.

Therefore, the receiver which can provide high capacity will perform the spectrum sensing early and have high probability for forwarding.

The estimation of capacity can be determined by Shannon's theorem. Let P_{sw}^j represent the probability of a particular node, such as node i, switching transmission to the $j_t h$ cognitive channel, which is given by

$$P_{sw}^j = P_b^j \left(1 - P_d^j\right) + P_i^j \left(1 - P_f^j\right) \tag{5}$$

We suppose that the upper bound of allowed power for transmission over the jth channel is $\mathcal{P}_{max}^j$. Then, the maximum capacity of channel j is given by

$$C_{\mathrm{x},y}^j = W_{x,y}^j \log_2 \left(1 + \frac{\mathcal{P}_{\max}^j \left|h_{x,y}^j\right|^2}{\delta^2}\right) \tag{6}$$

where $W_{x,y}^j$ denotes the probable bandwidth of the j^{th} channel allocated to CR users x and y for transmission. δ^2 and $h_{x,y}^j$ denote the power of noise and the channel coefficient, respectively.

The probable bandwidth really gives the available bandwidth as a result of CR users switching to a channel with certain probability given by (5). Then, $W_{x,y}^j$ can be obtained as follows.

$$W_{x,y}^j = P_{sw}^j B_j \tag{7}$$

where B_j is the bandwidth of the j^{th} channel.

Furthermore, the S will retransmit the packet, if all the involving nodes fail to deliver the preamble for forwarding in a particular time window. Node S can achieve this through immediately expiring the contention window (T_{cw}) after performing spectrum sensing.

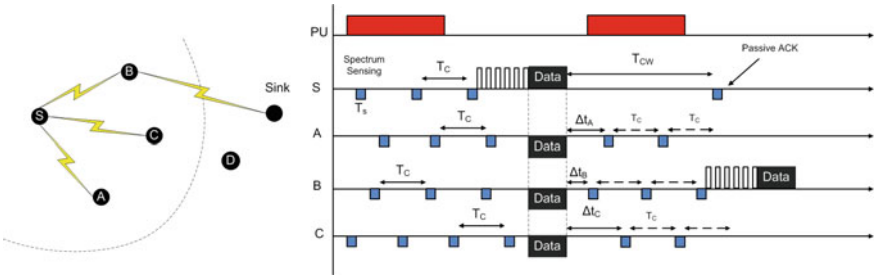


Fig. 1. Illustration of the timeline for a single hop in HCR-MAC

3 Performance Evaluation

The single hop performance of HCR-MAC will be evaluated in this section.The assumed topology of this protocol is shown as in Fig. 2. We assume that 16 PU transceivers are in the square region, which is with side 1200 meters. The smart meters are considered to follow the Poisson distribution in the whole region with a constant density. Without loss of generality, it is assumed that the RPL is used for simulation. We assume that the communication coverage is of each smart meter is within 40 m. In Table 1, other simulation parameters are shown. In addition, we compare the simulation results with the protocol in literature [7] and the CRB-MAC protocol in literature [6] in a cognitive enabled AMI network.

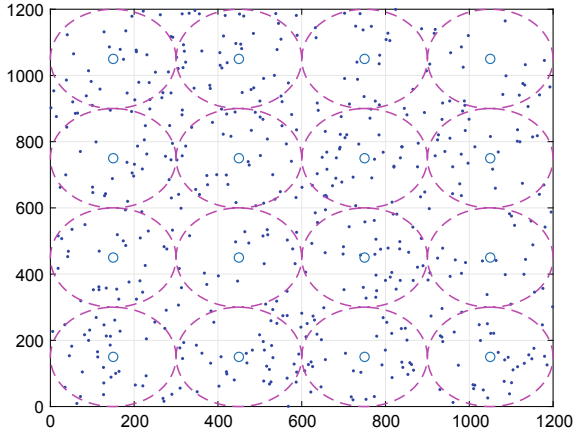


Fig. 2. Sample simulated topology. Smart meters distribute in Poisson distribution. The hollow circle denotes the PU location. The dotted circles denote the communication range of PU transceiver

First, we evaluate the average capacity of each channel when reaching their maximum allowable transmission power. The numerical results are illustrated in

Table 1. Parameters of simulation configuration

Parameter	Value
Path loss model	$127.1 + 36.6\log_{10}(r)$, r in km
Standard deviation of shadowing	7 dB
Detection probability threshold(P_d)	0.85
Probability of false alarm (P_f)	0.15
Channel bandwidth	200 kHz
PU received SNR (γ)	−16 dB
Busy state parameter of PU (μ_{ON})	[2, 8]
Idle state parameter of PU (μ_{OFF})	2
Maximum interference ratio (IR_{max})	0.26

Fig. 3. In each channel, the allowable maximum transmission power is randomly set to 20 dBm and 30 dBm. At each CR sensor node, 30 dBm is selected as the constraint of power. As Fig. 3 shown, the average capacity decreases with the decrease of the maximum transmission power constraint. After transmission power reaches the maximum allowed value, the corresponding average capacity peaks at the saturation point.

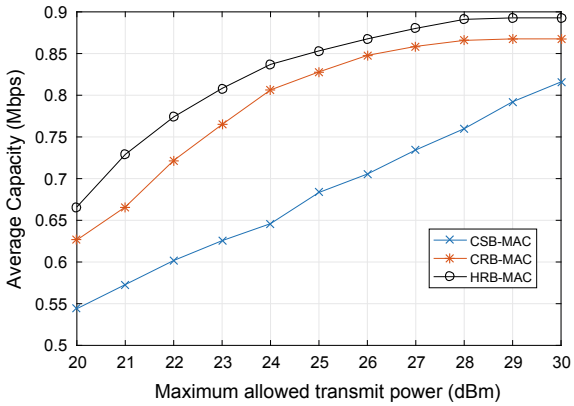


Fig. 3. Average capacity against bit error rate

Next, we evaluate the HCR-MAC with its reliability performance. We use PDR as the metric of reliability. The PDR performance against the bit error rate (BER) can be found in Fig. 4, where we can find that HRB-MAC and CRB-MAC perform better in reliability. This is mainly in that owing to the receiver-based mechanism, there is an increasing number of receiver nodes involved to forward the data. In conclusion, the HRB-MAC performs robust to channel quality fluctuate and has high reliability.

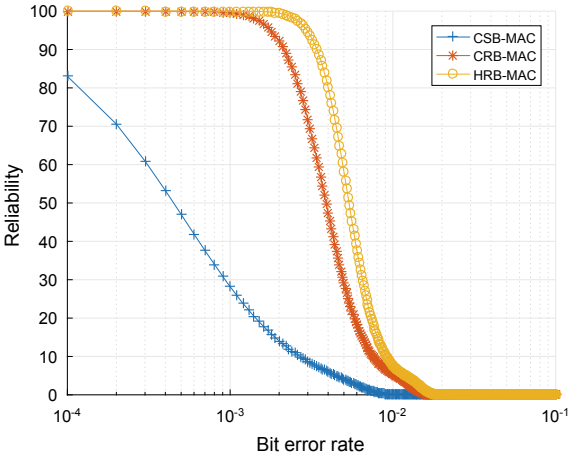


Fig. 4. PDR against BER

Last but not the least, Fig. 5 shows the delay performance against BER. Since delay depends on the times of retransmissions, HCR-MAC outperforms CRB-MAC and CSB-MAC considering delay performance due to fewer retransmissions. The delay performance decreases with the increase of receivers owing to higher proportion successful transmission and eventually reaches the climax as the upper limit of retransmission times is reached. It should be noted that with the same amount of receivers, a higher PU activity further aggravates the delay performance because of more frequent spectrum sensing to detect a channel vacancy.

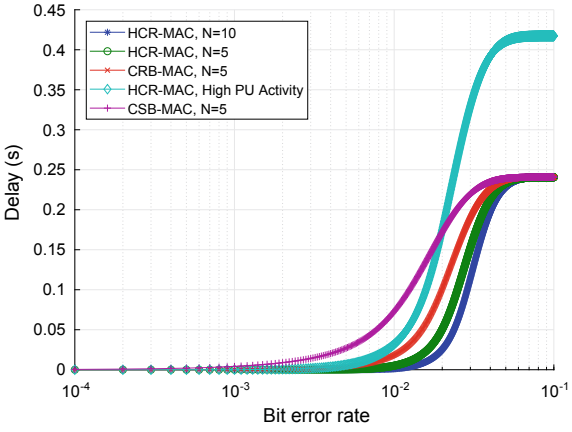


Fig. 5. Delay performance against BER

4 Concluding Remarks

In our work, a novel MAC protocol is proposed for CR-AMI networks, namely, HCR-MAC. HCR-MAC employs a capacity-based auction mechanism for forwarding competition, wherein the receiver with the best capacity will forward the data packet. Moreover, a preamble sampling approach is utilized to meet the requirements for high efficiency and reliability in smart grid. System level simulation results demonstrate that HCR-MAC performs less retransmissions in lossy wireless environments and hence improve the throughput and delay performance. Moreover, this protocol can supply more reliable data transmission through receiver competition for forwarding, which is due to the nature of a receiver-based protocol. Overall, the proposed HCR-MAC proves to be a potential solution for CR-AMI networks in fulfilling the expectation of smart grid.

References

1. Yang, Z., Han, R., Chen, Y., Wang, X.: Green-RPL: an energy-efficient protocol for cognitive radio enabled AMI network in smart grid. Access IEEE **14**(3), 799–821 (2018)
2. Li, P., Wu, M., Zhao, M., Liao, W.: A quality of serviceCaware preemptive tidal flow queuing model for wireless multimedia sensor networks in the smart grid environment. Int. J. Distrib. Sens. Netw. **13**(7), 1–18 (2017)
3. Yang, Z., Ping, S., Sun, H., Aghvami, A.H.: CRB-RPL: a receiver-based routing protocol for communications in cognitive radio enabled smart grid. IEEE Trans. Vehicular Technol. **66**(7), 5985–5994 (2017)
4. Cao, N., Chen, Y., Yang, Z.: Secrecy outage probability with randomly moving interferers in Nakagami-m fading, p. 1. Wireless Communications Letter, IEEE (2018)
5. Yang, Z., Ping, S., Nallanathan, A., Zhang, L.: An energy-efficient routing protocol for cognitive radio enabled AMI networks in smart grid. In: IEEE 83rd Vehicular Technology Conference (VTC Spring), pp. 1–5 (2016)
6. Aijaz, L.A., Ping, S., Akhavan, M.R., Aghvami, A.-H.: Crb-mac: a receiver-based mac protocol for cognitive radio rquipped smart grid sensor networks. IEEE Sens. J. **14**(12), 4325–4333 (2014)
7. Watteyne, T., Bachir, A., Dohler, M., Barthel, D., Auge-Blum, I.: 1-hopMAC: an energy-efficient MAC protocol for avoiding 1-hop neighborhood knowledge. In: IEEE International Conference on Sensor and Ad Hoc Communications and Networks (SECON), pp. 639–644 (2006)

Evaluation Method for Node Importance of Communication Network Based on Complex Network Analysis

JunYang Zhao(✉), Xiangwei Liu, and Jianpeng Guo

National University of Defense Technology Institute, Hefei 230037, China
2602905422@qq.com

Abstract. The time variant communication network is divided into a number of time slice communication networks by the time slice. We calculate the degree of concussion of node load redistribution to the nearest neighbor node, and regard it as the importance of nodes in the time slice communication network. Because of the different importance of each time slice, the concept and calculation method of the importance of the time slice communication network are put forward. Finally, the importance of the communication node in the whole communication period is the result of the importance of the node and the importance of the network.

Keywords: Complex network · Communication network · Node importance

1 Introduction

The communication network node shows different importance in the network because of its function, network location and organizational relationship [1–3]. The existing methods of measuring the importance of nodes can be divided into two categories. The first is to evaluate the centrality of the nodes in the network topology; two is to evaluate the decline of the network performance by removing a node [4–7]. Some research results show that the importance of network nodes is also affected by the propagation mechanism. Time-varying characteristics of communication networks, the access and exit of nodes will change the network topology. So the time of node access in the communication period also affects the node importance. Based on the theory of complex networks and taking into account the time domain characteristics of communication networks, this paper presents a method for calculating the importance of nodes.

2 Evaluation Method of Node Importance Degree in Communication Networks

The link connection of communication network can be represented by digraph. Such as $G = (V, E)$. In the formula, V is a node, representing the information receiving and processing terminal, E as the edge, representing the connection of each terminal, and

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 342–349, 2019.
https://doi.org/10.1007/978-981-13-6264-4_42

stipulate that there is only one edge between any two nodes, and there is no self-ring. Communication network is a dynamically changing network. Node access and exit change the topology of the network. Therefore, when analyzing the node importance, starting with the network topology, this paper studies the importance of the nodes of the communication network which is stable in a certain period of time, and then evaluates the importance of the communication network in the whole communication period. The mathematical model for evaluating the importance of communication network nodes is

$$g_i = \sum_{k=1}^{n} h_{ki} \cdot w_k \tag{1}$$

In the case, where g_i is the importance of node i in the entire communication period; h_{ki} is the importance of nodes in the communication network at the i sub-period; w_k is the importance of the communication network of the i sub-period in the whole communication period; n is the number of sub-periods after the division of the communication period.

3 Complex Network Node Importance Evaluation Model

3.1 Node State

The node state indicates the ability of nodes to accommodate the load transfer. It is determined by two factors: one is the load remaining capacity of the node itself, and the two is the residual capacity of the load of the adjacent nodes. From the above description process, it can be divided into two stages. First considering the remaining capacity of the node, it is assumed that the real time load of node i is Q_i, and the capacity of the node is C_i, then the remaining capacity R_i of the node can be expressed as

$$R_i = \begin{cases} C_i - Q_i, & C_i \leq Q_i \\ 0, & C_i > Q_i \end{cases} \tag{2}$$

$R_i = 0$ represents that the node has failed. The larger the R_i, the stronger the node's processing power. It should have an advantage in load redistribution.

Considering the influence of the load residual capacity of adjacent nodes, we propose a more precise way to measure the node state of node load residual capacity, such as formula (3). In the formula, Z_i represents node state; R_i R_α and R_α represents node remaining capacity; Φ_i represents the adjacency node aggregate of node i; λ is the weight parameter. By adjusting the value of λ, the proportion of node state in load redistribution can be adjusted. When λ is 0, the formula (3) indicates the uniform redistribution of local load.

$$Z_i = \begin{cases} \left[\min\left(R_i, \sum\limits_{\alpha\in\Phi_i} R_\alpha\right)\right]^{\lambda}, & \alpha \notin \emptyset \\ R_i^{\lambda}, & \alpha \in \emptyset \end{cases} \tag{3}$$

3.2 Load Redistribution

After the node fails, the load on the node will be transferred to the adjacent nodes to ensure the effective operation of the entire network. The whole network will update the load due to the failure of nodes. The update rule is $Q_i' = Q_i + \Delta Q_i, i \in \Phi_f$. In the formula, Q_i is the load before updating; Q_i' is the load after updating; ΔQ_i is the increment of update; Φ_f is the all adjacent nodes of invalid node f. Assume that the load of failed node f is Q_f, and the node state of adjacent nodes is Z_i. The formula for the calculation of ΔQ_i is

$$\Delta Q_i = Q_f \frac{Z_i}{\sum\limits_{\alpha\in\Phi_f} Z_\alpha}, i \in \Phi_f \tag{4}$$

It is an improvement to the classical neighborhood redistribution method, which is related to node state but not to node load.

3.3 Evaluation of Node Importance

After the node f fault, because its own load is transferred to the neighbor node, the degree of the transfer of the neighbor nodes to the failure state is used to measure the ability of the node f to cause the cascading failures of the network, that is the importance of the node [9]. In order to avoid further cascading failures, the load of the nodes after reassignment is limited by its capacity. The capacity of the node is $C_i = (1+\eta)Q_{0i}$. In the formula the constant η is the maximum tolerance of the network, which represents the ability to handle extra load; Q_{0i} is the initial load of nodes. Therefore, the redistribution load will satisfy the limitation of formula (5).

$$\Delta Q_i = Q_f \frac{Z_i}{\sum\limits_{\alpha\in\Phi_f} Z_\alpha} < \eta Q_{0i}, i \in \Phi_f \tag{5}$$

Therefore, we can use $Q_f Z_i / [\eta Q_{0i}(\sum\limits_{\alpha\in\Phi_f} Z_\alpha)]$ to indicate the load concussion degree of neighbor nodes after node f failure. Therefore, the importance degree of node f can be defined by the average failure degree of neighbor nodes, as shown in formula (6).

$$h_f = \sum_{i=\Phi_f} \frac{Q_f Z_i}{\eta Q_{0i}(\sum\limits_{\alpha\in\Phi_f} Z_\alpha)} / |\Phi_f| \tag{6}$$

In the formula, $|\Phi_f|$ is the number of all adjacent nodes of the fault node.

4 Time Slice Communication Network Importance Evaluation Model

In the time domain, the importance of a communication network depends on the following two points. One is the importance of the time interval of the network in the whole period of communication, such as the communication network that initiates the attack stage in the military communications, which is more important than the communication network in the reconnaissance stage. Two is the relative time length of the network, such as the frequently active nodes in the communication network have more information processing tasks, and these nodes constitute a communication network with high importance degree. However, nodes with low activity are usually spare nodes or remote nodes, and their importance is relatively low. Based on the above principles, we set up a communication network importance evaluation model. We divide the communication period into a number of nonoverlapping time slices [8], and the topology of the communication network in each time slice is unchanged. So the communication network of time-varying is decomposed into a static communication network based on multiple time slices. Each time slice corresponds to a static communication network structure, which is different from the static communication network structure corresponding to the previous time slice or the later time slice. Then we calculate the importance of the communication network based on the importance of time slice in the communication period.

According to the above ideas, first, U(the whole communication time segment) is divided into L nonoverlapping time slices, which is expressed as $\{[b_1, e_1]; [b_2, e_2]; \ldots; [b_L, e_L]\}$. Then the whole communication period is divided into N subsections by the communication decision maker, which is expressed as $\{[T_1, T_2]; [T_2, T_3]; \ldots\ldots; [T_N, T_{N+1}]\}$. The decision maker decides the importance of each subsections based on the method of expert scoring according to the location and task of the subsections in the whole communication time, which is expressed as $\{\theta_1; \theta_2; \ldots\ldots; \theta_N\}$, in the formula $\sum_{i=1}^{N} \theta_i = 1$.The segmentation of the time slice is shown in Fig. 1. The importance of the time slice k is

$$r_k = [\frac{T_{p+1} - b_k}{e_k - b_k} \cdot \theta_p + \sum_{i=2}^{X-1} (\frac{T_{p+i} - T_{p+i-1}}{e_k - b_k} \cdot \theta_{p+i-1}) + \frac{e_k - T_{p+X-1}}{e_k - b_k} \cdot \theta_{p+X-1}] \cdot \frac{e_k - b_k}{\sum_{i=1}^{L} (e_i - b_i)} \tag{7}$$

In the formula $T_p \le b_k < T_{p+1}, T_{p+X-1} \le e_k < T_{p+X}$, after normalizing the formula (7), we get that the importance of communication network corresponding to time slice k is $w_k = \frac{r_k}{\sum_{i=1}^{L} r_k}$.

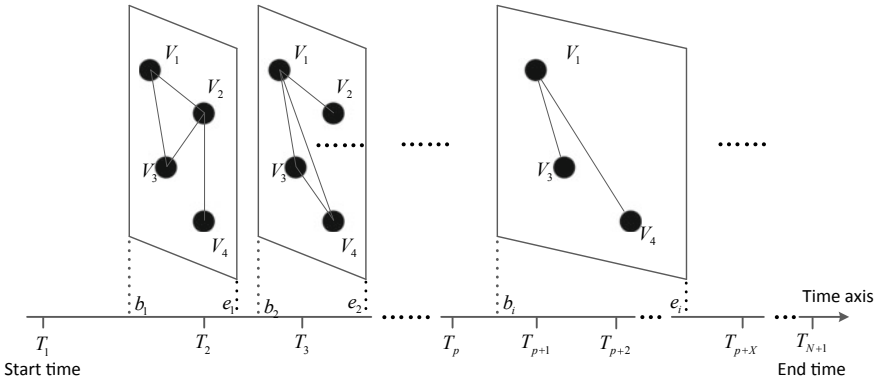


Fig. 1. Schematic diagram of communication networks corresponding to time slices

5 Example Analysis

5.1 Construction of Communication Network

The following two characteristics are mainly considered when constructing the communication network needed for instance analysis. One is the time-varying characteristics of the network. New nodes continue to join the network, and the old nodes quit the network. Two, new nodes often choose nodes with larger degree to establish connections, and most of the nodes that exit from the network are nodes with lower degree. Therefore, a time-varying stochastic network model is proposed in this paper. The algorithm of the model is as follows:

Step 1 the initial setup: First, a random network with j_0 nodes is established, and each node is numbered according to 1 to j_0.

Step 2 increase of nodes: A new node O is added every B seconds, and it is numbered $j_0 + 1$. The node O connects m nodes of the original network. In the above content $m \leq j_0$, $B \sim N(B_0, \sigma_1^2)$, $m \sim N(m_0, \sigma_2^2)$.

Step 3 preferential connection: The selection of connection nodes is based on the degree value priority principle. The probability of selecting the old i node as the new edge to connect node is $\Pi_i = \frac{D(i)}{\sum_j D(j)}$, in the formula, $D(\cdot)$ is a node degree value.

Step 4 node withdrawal: The node with the smallest degree is out of the network. If there are multiple nodes with the least degree, the one is chosen randomly. After the node is withdrawn, all links connected to it are disconnected.

Step 5 value assignment of communication subsections: In real cases, communication sub-segments should be partitioned and valued by decision makers according to subjective judgement and objective reality. Usually, subsection near the middle time takes more information transmission tasks, and its importance is higher. In simulation, in order to simplify the process, the whole communication time is evenly divided into N sub-segments, which is expressed as $\{[T_1, T_2]; [T_2, T_3]; \ldots\ldots; [T_N, T_{N+1}]\}$. The

importance of each subsection is given by the Gauss distribution. The importance of the i subsection is

$$\theta_i = \frac{1}{\sqrt{2\pi}\sigma_3}\exp\{-\frac{[(T_i + T_{i+1})/2 - (T_1 + T_{N+1})/2]^2}{2\sigma_3^2}\} \tag{8}$$

In the formula, $(T_1 + T_{N+1})/2$ is the mean of Gauss distribution, and σ_3^2 is the variance.

5.2 Simulation Analysis

The parameters in the above time-varying stochastic network model are as follows: $j_0 = 200$, $S_0 = 3$, $B_0 = 100s$, $m_0 = 3$, $\sigma_1 = 10s$, $\sigma_2 = 1$, $\sigma_3 = 60s$, $N = 6$, $T_1 = 0s$, $T_{N+1} = 3600s$. The initial load for each node is $Q_{0i} = ak_i^{\gamma}$, k_i is the node's degree; a and γ are adjustable parameters, controlling the initial load. For simplified calculation, among all the nodes, the values of a and γ are 1 and η is 0.07. Based on the node importance evaluation method, the calculation results are shown in Table 1. Table 1 shows the top 30 nodes of the rank of importance.

Table 1. Node importance

Rank of importance	Node numbering	Value of importance	Rank of importance	Node numbering	Value of importance
1	18	0.7791	16	42	0.4455
2	25	0.7791	17	14	0.4325
3	150	0.7791	18	131	0.4225
4	180	0.7791	19	81	0.4126
5	83	0.7791	20	216	0.4111
6	197	0.7101	21	51	0.4029
7	140	0.7101	22	137	0.3471
8	89	0.6901	23	185	0.3321
9	85	0.6240	24	145	0.3296
10	174	0.6008	25	144	0.3224
11	193	0.5701	26	95	0.3223
12	155	0.5423	27	201	0.3223
13	178	0.4631	28	90	0.3223
14	8	0.4602	29	43	0.3201
15	71	0.4600	30	48	0.3110

According to node importance, from high to low, the node is invalid from the network. And failure influence degree ($\Omega(i)$) is used as an index to evaluate the degree of influence on the whole communication network after failure of nodes. The definition of the influence degree of node i failure is shown in formula (9). Among, $\psi_{i,j}/200$ represents the ratio of nodes cascading failure after node i fails on the network of time

slice j. The calculation results of the degree of node failure influence are shown in Fig. 2.

$$\Omega(i) = \sum_{j=1}^{L} [\frac{\psi_{i,j}}{200} \cdot r_j \cdot \frac{e_j - b_j}{\sum_{k=1}^{L} (e_k - b_k)}] \tag{9}$$

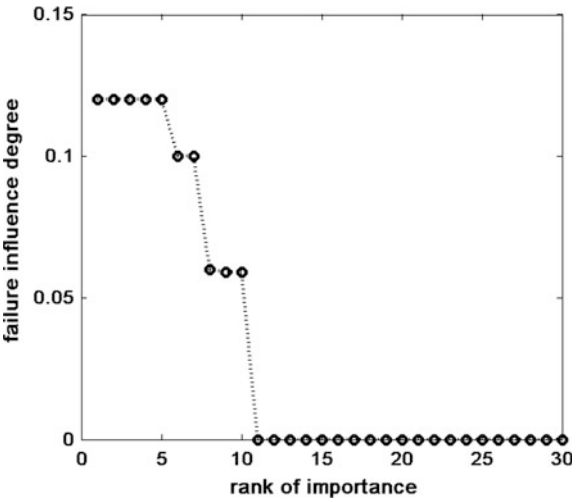


Fig. 2. Failure effect of node in communication network

The abscissa of Fig. 2 is the importance ranking of failure nodes. The ordinate represents the failure influence degree. From the graph, we can see that the impact of node failure with higher importance is significantly greater than that of nodes with lower importance. For this communication network, the 10 nodes with the top rank have strong failure effects, while the nodes with lower importance degree have weaker failure influence. It is proved that this node importance evaluation method is suitable for this communication network.

6 Summary

In this paper, the importance degree evaluation method of the communication network node is proposed. On the basis of traditional evaluation method, we consider the influence of communication time on node importance. At the same time, we consider the importance of nodes in the network topology, and the importance of nodes in network propagation. It is relatively true to reflect the importance of nodes, which helps to locate the key nodes in communication networks. But in the evaluation model, the function and business of each node have not been distinguished, which will be the focus of next research.

References

1. Jing xing, L., Shao dong, C., et al.: Study on node importance of complex network based military command control networks. In: Proceedings of International Conference on Machine Learning and Cybernetics, pp. 920–923. IEEE Press, Xian (2012)
2. Tao, Z., Guo cen, C., Rong xiao, G., et al.: Research on centrality model and evaluation of networked command and control . J. Syst. Simul. **22**(1), 201–204 (2010)
3. Xuan, Z., Feng ming, Z.: Evaluating complex network functional robustness by node efficiency. Acta Phys. Sinical **61**(19), 1–7 (2012). (in Chinese)
4. Jian guo, L., Zhuo ming, R., Qiang, G., et al.: Node importance ranking of complex networks. Acta Physica Sinica **62**(17), 1–10 (2013). (in Chinese)
5. Klemm, K., et al.: A measure of individual role in collective dynamics. Sci. Rep. **292**(2), 1–8 (2012)
6. Aral, S., Walker, D.: Identifying influential and susceptible members of social networks. Science **337**(6092), 337–341 (2012)
7. Guang quan, C., Yong Zhong, L., Ming xing, Z., Jin cai, H.: Node importance evaluation and network vulnerability analysis on complex network. J. Natl Univ. Def. Technol. **39**(1), 120–127 (2017)
8. Ming zhi, Z., Kai, L., Xi, W.: Study on key nodes analysis method of space information network. J. Syst. Simul. **27**(6), 1235–1239 (2015)
9. Dong li, D., Ren jun, Z.: Evolution mechanism of node importance based on the information about cascading failures in complex networks. Acta Physica Sinica **63**(6), 381–389 (2014). (in Chinese)

A Data Fusion Scheme for Wireless Sensor Networks Using Clustering and Prediction

Xiulan Yu, Hongyu Li(✉), Chenquan Gan, and Zufan Zhang

Chongqing Key Labs of Mobile Communications Technology,
Chongqing University of Posts and Telecommunications, Chongqing 400065, China
849645620@qq.com

Abstract. This paper intends to reduce the communication cost, while ensuring data prediction accuracy and data transmission efficiency in wireless sensor networks (WSNs). A data fusion scheme using clustering and prediction algorithms is proposed. Initially, nodes are clustered by using historical data, and then they are linked based on the actual geographical distance. Next, during the data fusion process, the base station and sensor nodes both use the online recurrent extreme learning machine (OR-ELM) to predict the future sensing data, which can guarantee that the data sequence in the base station and sensor nodes are synchronous. If the prediction fails, data will be transmitted to other nodes in the link and forwarded. Finally, experimental results reveal that the proposed data fusion scheme not only can effectively predict the sensor data, but also can reduce spatial and temporal redundant transmissions with low computational cost.

Keywords: Wireless sensor network · Data fusion · Clustering · Time-series prediction

1 Introduction

WSNs have been successfully applied in many fields [1], such as surveillance, environmental monitoring, smart city, and health care, so a large number of sensor nodes will perceive the information of various monitored objects and send this information to the base station and end-users [2]. Meanwhile, the requirements of data collection and data fusion are becoming increasingly higher.

In big data WSNs, the sensed data have a massive similarity and redundancy [3], and the spatial and temporal correlation of data can lead to large amounts of redundant transmissions and the waste of node energy [4]. As a result, clustered WSNs are the prominent network architecture for data reduction strategies [5]. In particular, each cluster uses a specific data compression strategy to perform a

Hongyu Li, M.S., Chongqing University of Posts and Telecommunications. His current main research interest includes: wireless sensor networks and extreme learning machine.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 350–358, 2019.
https://doi.org/10.1007/978-981-13-6264-4_43

series of compression operations on the cluster head [6]. Additionally, a number of compression methods were proposed, such as compressed sensing (CS) [7]. Due to the limited storage space and computation capacity of sensors, CS is inefficient for data collection at cluster head nodes. Then, data prediction was discussed as the more efficient way to obtain data reduction in WSNs [8]. Autoregressive (AR) model was used to calculate the estimation of future perceived data. However, the shortcoming of this model is that communication cost would be very high when the error threshold was set to a small value. Afterward, the GM-OP-ELM scheme, which combined gray model (GM) and optimally pruned extreme learning machine (OP-ELM) was proposed [9], and [10] considered the GM-KRLS scheme, which consisted of GM and kernel recursive least squares (KRLS). It is easy to see that both of these algorithms are based on grey model prediction. Online sequential extreme learning machine (OS-ELM) is known to be a viable solution to time-series prediction [11]. And recently, a new variant of OS-ELM, called online recurrent extreme learning machine (OR-ELM), was proposed [12].

Inspired by the above-mentioned work, a data fusion scheme for WSNs using clustering and OR-ELM prediction algorithms is proposed. First, nodes are clustered by considering both the distance of nodes and trend similarity between the node data series, and strong links between them are established. Second, OR-ELM prediction algorithm is used to reduce sampling data points. Third, prediction-failed nodes find similar nodes through different links, and only one node sends the data in the link. Finally, experimental results show that the proposed data fusion scheme not only can effectively predict the sensor data, but also can reduce spatial and temporal redundant transmissions with low computational cost.

The organization of the rest of this paper is as follows: Sect. 2 introduces the proposed scheme. In Sect. 3, some experimental results are explained. Finally, Sect. 4 summarizes this work.

2 Description of the Proposed Data Fusion Scheme

As shown in Fig. 1, the main steps of the proposed scheme are as follows:

Step 1: Through the historical data sequence, the similarity of nodes is calculated. Then, a similar node cluster is obtained using the clustering algorithm.

The sensing objects include temperature, humidity, illumination, gas concentration. So the data sequence $\mathbf{S}_i$ can be represented as an array:

$$\boldsymbol{S}_i = \left[\, a_i\,(1) \cdots a_i\,(t) \;\; b_i\,(1) \cdots \; b_i\,(t) \;\; \cdots n_i\,(1) \;\; \cdots n_i\,(t) \, \right], \tag{1}$$

where $a_i\,(t)$, $b_i\,(t)$, and $n_i\,(t)$ represent the different sampled data of node i at time slot t. In order to get a better clustering result, all data with the same attributes are normalized to the range [0,1] according to

$$s_t = (s_t - s_{\min}) \,/\, (s_{\max} - s_{\min}), \tag{2}$$

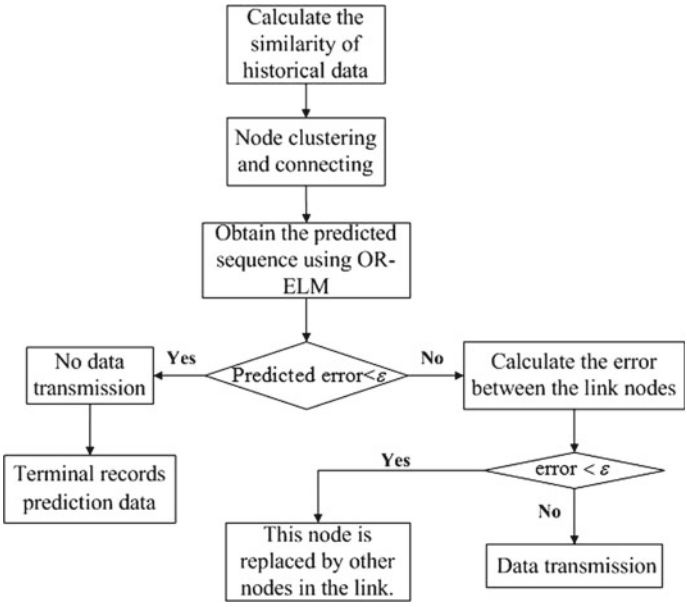


Fig. 1. The flowchart of the proposed scheme.

where $s_{\min}$ and $s_{\max}$ are the maximum and minimum of the sampled data s_t during the period $[0, t]$. Then, the Euclidean distance between two nodes can be obtained as follows:

$$\Phi(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}, \quad (3)$$

where $\Phi(x, y)$ is similarity between nodes x and y, x_i and y_i are the corrosponding sample data. Next, the k-means method [13] is improved and applied here based on the following rules:

(R1) Define the number of randomly selected k groups (clusters).
(R2) Calculate the similarity between each point.
(R3) Classify the points where the similarity is less than a threshold into a cluster.
(R4) Update the cluster center and repeat second and third steps until the center is not changed.

Step 2: Nodes are linked based on the actual geographical distance and the strong links are obtained. If the actual geographic distance of two nodes in cluster is less than a threshold, then they will be linked. As shown in Fig. 2, the solid line are strong links, and the dotted line are weak links.

Step 3: Construct and train the OR-ELM using q training samples and predict future time series. Figure 3 shows the network topology of the OR-ELM.

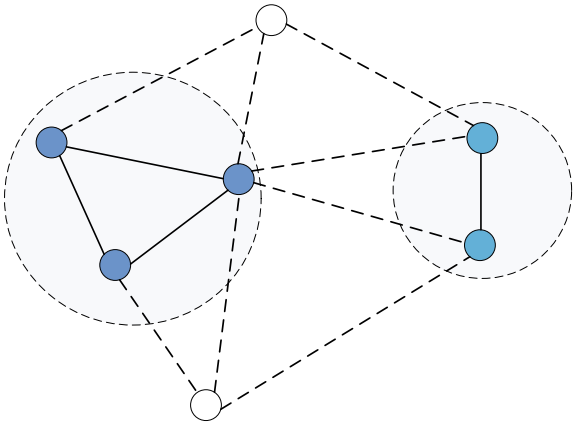


Fig. 2. Link state.

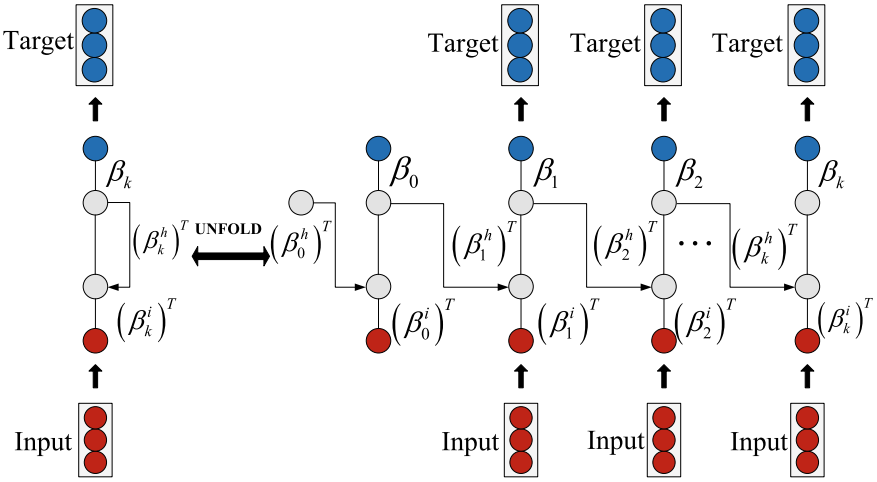


Fig. 3. The network structure of the OR-ELM.

ELM is a learning algorithm for training an single-hidden-layer feedforward network (SLFN). For N training samples (x_j, t_j), the output of the ELM with L hidden nodes can be denoted by

$$\sum_{i=1}^{L} \beta_i g\left(\omega_i \cdot x_j + b_i\right) = t_j, j = 1, 2, \cdots, N, \tag{4}$$

where β_i is the output weight, ω_i and b_i severally denote input weights and bias values of hidden nodes, $g(\cdot)$ is a nonlinear activation function. Therefore, it can be further expressed compactly as a matrix:

$$\mathbf{H}\beta = \mathbf{T}, \tag{5}$$

where $\mathbf{H}$ denotes the hidden-layer output matrix of the $SLFN$ and the output weights is given as follows:

$$\beta = \mathbf{H}^{\dagger}\mathbf{T}, \mathbf{H}^{\dagger} = \left(\mathbf{H}^{T}\mathbf{H} + \frac{\mathbf{I}}{C}\right)^{-1}\mathbf{H}^{T}, \tag{6}$$

where $\mathbf{H}^{\dagger}$ is the Moore–Penrose generalized inverse of matrix C and $\mathbf{H}$ is a regularization constant. The output weight β_0 for an initial training dataset with N_0 training samples is calculated as

$$\beta_0 = \mathbf{P}_0\mathbf{H}_0\mathbf{T}_0, \quad \mathbf{P}_0 = \left(\mathbf{H}_0^{T}\mathbf{H}_0 + \frac{\mathbf{I}}{C}\right)^{-1}. \tag{7}$$

During the online sequential learning phase, whenever the new input data of the N_{k+1} training sample arrives, the output weight is updated:

$$\beta_{k+1} = \beta_k + \mathbf{P}_{k+1}\mathbf{H}_{k+1}^{T}\left(\mathbf{T}_{k+1} - \mathbf{H}_{k+1}\beta_k\right), \tag{8}$$

$$\mathbf{P}_{k+1} = \mathbf{P}_k - \mathbf{P}_k\mathbf{H}_{k+1}^{T}\left(\mathbf{I} + \mathbf{H}_{k+1}\mathbf{P}_k\mathbf{H}_{k+1}^{T}\right)^{-1}\mathbf{H}_{k+1}\mathbf{P}_k, \tag{9}$$

where $k+1$ is the $(k+1)$th chunk of input data with k increasing from zero, and H_{k+1} indicates the hidden-layer output for the $(k+1)$th chunk of input data. The input sample $x\,(k+1) \in R^{n\times 1}$ is propagated to the hidden layer so hidden-layer output matrix $\mathbf{H}_{k+1}^{i}$ can be calculated as follows:

$$\mathbf{H}_{k+1}^{i} = g\left(norm\left(\mathbf{X}_{k+1}^{i}x\,(k+1)\right)\right). \tag{10}$$

Note that the *norm* function is added before the nonlinear activation as a layer normalization procedure to prevent the problem of internal covariate shift. Then, one can calculate output weight β_{k+1}^{i} using RLS:

$$\beta_{k+1}^{i} = \beta_k^{i} + \mathbf{P}_{k+1}^{i}{\mathbf{H}_{k+1}^{i}}^{T}\left(x\,(k+1) - \mathbf{H}_{k+1}^{i}\beta_k^{i}\right), \tag{11}$$

$$\mathbf{P}_{k+1}^{i} = \frac{1}{\lambda}\mathbf{P}_k^{i} - \mathbf{P}_k^{i}{\mathbf{H}_{k+1}^{i}}^{T}\left(\lambda^2 + \lambda\mathbf{H}_{k+1}^{i}P_k^{i}{\mathbf{H}_{k+1}^{i}}^{T}\right)^{-1}\mathbf{H}_{k+1}^{i}\mathbf{P}_k^{i}. \tag{12}$$

Set the forgetting factor $\lambda \in (0,1]$. The transpose of β_{k+1}^{i} serves as the input weight of OR-ELM $\mathbf{X}_{k+1}$:

$$\mathbf{X}_{k+1} = {\beta_{k+1}^{i}}^{T}. \tag{13}$$

Hidden-layer output matrix $\mathbf{H}_{k+1}^{h}$ is given by

$$\mathbf{H}_{k+1}^{h} = g\left(norm\left(\mathbf{X}_{k+1}^{h}\mathbf{H}_k\right)\right). \tag{14}$$

Then, using RLS calculates output weight β_{k+1}^{h} and the transpose of β_{k+1}^{h} is used as the hidden weight of OR-ELM $\mathbf{Y}_{k+1}$:

$$\mathbf{Y}_{k+1} = {\beta_{k+1}^{h}}^{T}. \tag{15}$$

Now, for the $(k+1)$th input $x(k+1)$, the OR-ELM hidden-layer output matrix $\mathbf{H}_{k+1}$ is calculated as follows:

$$\mathbf{H}_{k+1} = g\left(norm\left(\mathbf{X}_{k+1}x\left(k+1\right) + \mathbf{Y}_{k+1}\mathbf{H}_k\right)\right). \tag{16}$$

Step 4: The base station and sensor nodes both use the OR-ELM to predict the future perceived data, which can guarantee that the data sequence in the sensor nodes and base station are synchronous. If the error between them is below the threshold, then it is unnecessary for the sensor node to transmit data to the base station. When the data prediction target is achieved, the data need to be saved. At the same time, the terminal treats the predicted data as perceived data for the current period.

Step 5: If the prediction fails, data will be transmitted to other nodes in the link and forwarded. Figure 4 displays the nodes connection diagram, and the blue nodes represent nodes that need to transmit data.

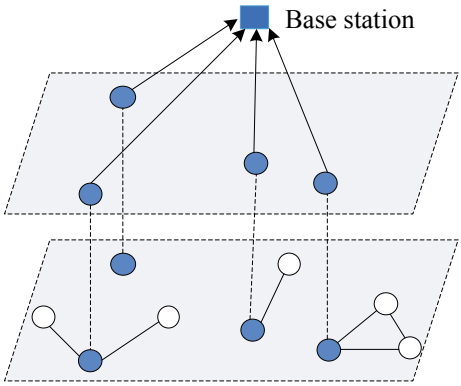


Fig. 4. Network topology.

3 Experimental Results

In order to evaluate the proposed scheme, some simulations are examined based on Intel Labs' data collected from Intel Labs during a 1-month period. The real data set is collected from 54 distributed sensors.

The OR-ELM input dimension is configured to 250 steps with a 250-step time delay. The output dimension of the OR-ELM is set to 1. That is, the proposed scheme will continue to accept sequences that take the past 250 steps as input, and as a target output, there are 5 steps of future values. Because 250 data sampling points for each node are used for clustering, the statistical data analysis is also performed for the data after 250 sampling points. Set the forgetting factor $\lambda = 0.96$, and $L = 25$ denotes the number of hidden nodes.

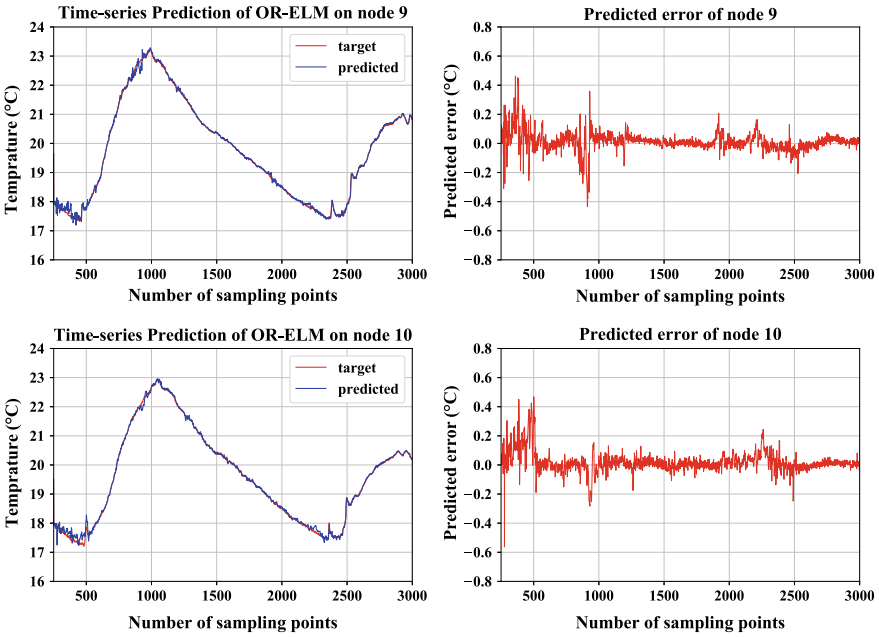


Fig. 5. Prediction results.

Two nodes are randomly selected to prove the performance of the algorithm. Figure 5 shows nodes 9 and 10 time-series prediction of OR-ELM on dataset. Comparing the prediction success rates for nodes 9 and 10 as the error threshold e-max changes from 0° to 0.22°. As shown in Fig. 6, when e-max = 0.1, the algorithm achieves about 90% successful prediction for all nodes. In other words, over 90% of the collected sensor data is not to be transmitted to the base station. With the increase of threshold, the algorithm gets a significant improvement in reducing communication. Clustering results prove that nodes 9 and 10

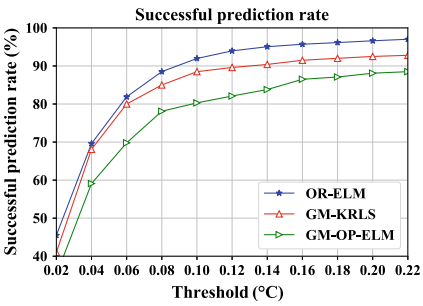


Fig. 6. The comparison of the prediction success rates.

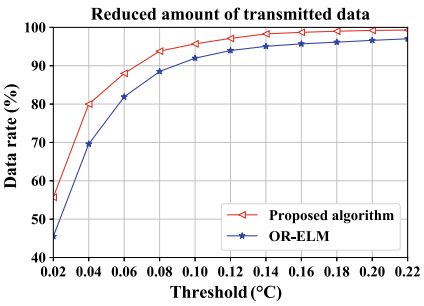


Fig. 7. The comparison of the reduced amount of transmitted data.

are strong links. The threshold is set to 0.1. So the successful rate of nodes 9 and 10 is 91.96% and 88.76%, respectively. The error of 65.5% of their real data is less than 0.1. The calculation yields that there are only 85 sample points that node 9 predicts failed and then is not similar to node 10. In other words, it senses the 2750 data, only 85 sampling data need to be sent to the terminal. There are also 135 sampling data that are predicted to fail. But they are similar to node 10 and can be replaced by values of node 10. Figure 7 shows the reduced rate of transmitted data. Although the results only report each node, the data reduction of the sensor nodes will result in a reduction of each cluster.

4 Conclusions

This paper has improved the performance of a data fusion scheme of a wireless sensor data transmission using clustering and prediction. The OR-ELM has shown higher accuracy than other algorithms, and highly correlated data between nodes have been clustered. Additionally, nodes can compare with connected nodes before transmission to further reduce the amount of data. Finally, experimental results have shown that the proposed scheme improved the prediction accuracy and reduced spatial and temporal redundant transmissions with low computational cost.

References

1. Durisic, M.P., Tafa, Z., Dimic, G., et al.: A survey of military applications of wireless sensor networks. Embedded Comput. pp. 196–199. IEEE (2012)
2. Tan, R., Xing, G., Liu, B., et al.: Exploiting data fusion to improve the coverage of wireless sensor networks. IEEE/ACM Trans. Netw. **20**(2), 450–462 (2012)
3. Zhang, C., Wang, B., Fang, S., et al.: Clustering algorithms for wireless sensor networks using spatial data correlation. In: International Conference on Information and Automation, pp. 53–58. IEEE (2008)
4. Vuran, M.C., Akyildiz, I.F.: Spatio-temporal correlation: theory and applications for wireless sensor networks. Comput. Netw. **45**(3), 245–259 (2004)
5. Arunraja, M., Malathi, V., Sakthivel, E.: Distributed similarity based clustering and compressed forwarding for wireless sensor networks. ISA Trans. **59**, 180–192 (2015)
6. Cheng, L., Guo, S., Wang, Y., Yang, Y.: Lifting wavelet compression based data aggregation in big data wireless sensor networks. In: IEEE, International Conference on Parallel and Distributed Systems, pp. 561–568. IEEE (2017)
7. Zheng, H., Yang, F., Tian, X., et al.: Data gathering with compressive sensing in wireless sensor networks: a random walk based approach. IEEE Trans. Parallel Distrib. Syst. **26**(1), 35–44 (2014)
8. Tulone, D., Madden, S.: PAQ: time series forecasting for approximate query answering in sensor networks. In: European Workshop on Wireless Sensor Networks, Vol. 3868, pp. 21–37. Springer, Berlin, Heidelberg (2006)
9. Luo, X., Chang, X.: A novel data fusion scheme using grey model and extreme learning machine in wireless sensor networks. Int. J. Control Autom. Syst. **13**(3), 539–546 (2015)

10. Luo, X., Zhang, D., Yang, L.T., et al.: A kernel machine-based secure data sensing and fusion scheme in wireless sensor networks for the cyber-physical systems. Future Gener. Comput. Syst. **61**, 85–96 (2016)
11. Liang, N., Huan, G., Saratchandran, P., Sundararajan, N.: A fast and accurate online sequential learning algorithm for feedforward networks. IEEE Trans. Neural Netw. **17**(6), 1411–1423 (2006)
12. Park, J.M., Kim, J.H.: Online recurrent extreme learning machine and its application to time-series prediction. In: International Joint Conference on Neural Networks, pp. 1983–1990. IEEE (2017)
13. Aznaoui, H., Raghay, S., Aziz, L.: New smart nodes distribution using kmeans approach to enhance routing in wsn. Indian J. Sci. Technol. **9**(46) (2016)

Joint Cooperation Set Selection and Resource Allocation of JT CoMP in Ultra Dense Network

Siyun Chen, Ye Wang(✉), Jia Yu, Yan Yan, and Niannian Wang

Communication Engineering Research Center, Harbin Institute of Technology, Shenzhen, Guangdong, China
{chensiyun,yanyan,wangniannian}@stu.hit.edu.cn
wangye.hitsz@qq.com,yujia_hitsz@hotmail.com

Abstract. Ultra Dense Network(UDN) is considered as a promising technology in the fifth-generation (5G) mobile communication system. However, some users will acquire extremely low sum rate as a consequence of strong inter-cell interference. Hence, Joint Transmission Coordinated Multi-Point(JT CoMP) is applied to mitigate the inter-cell interference. In this paper, to improve fairness of users, we optimize the user-centric selection of cooperation set, as well as considering the resource allocation to maximize the sum rate of the minimum user. Then we propose Load-aware Binary Genetic Algorithm(LBGA) and Resource Balancing-oriented RB Allocated Algorithm(RB-RBAA), respectively. Simulation results show that our algorithms can improve the sum rate of minimum user efficiently and guarantee the fairness of users.

Keywords: UDN · CoMP · Cooperation set selection · RB allocation

1 Introduction

Confronting with the increasing need for the network traffic, deploying large density of small Base Stations(BS) is considered to enhance the traffic of the network in 5G, which forms the Ultra Dense Network(UDN) [1]. Thanks to the smaller distances, the UDN will improve the spectral efficiency and Quality of Service(QoS) of users [2]. However, users located at edges of cells suffer from severer inter-cell interference in UDN, especially when the frequency reuse factor is 1. To deal with inter-cell interference, Coordinated Multi-point transmission (CoMP) is introduced by 3GPP in Release 9. It is proved to be helpful to enhance the Signal-to-Interference-plus-Noise Ratio(SINR) of cell edge users [3]. We focus on Joint Transmission CoMP(JT CoMP) and use "CoMP" to represent "JT CoMP" in the rest of this paper.

For the technology of CoMP, cooperation set selection of each user is crucial, since it significantly influences network performance. Cooperation set selection methods are widely attentioned in former literature. In [4], a specific user chooses

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 359–366, 2019.
https://doi.org/10.1007/978-981-13-6264-4_44

the BSs of size N which have the biggest Reference Signal Received Power (RSRP), which is called the N_{best} way. However, in a network with restricted resources, fairness of users is poor by this method. In [5], the cooperation set is determined by a threshold of SINR. The BSs whose RSRPs are beyond the threshold will become the candidates for cooperating set of a user. Nevertheless, since the condition of wireless channels is changing all the time, it is hard to determine the proper threshold of SINR. As a consequence, a tractable method pursuing fairness of users should be considered.

In a CoMP system, several BSs collaboratively serve a specific user. In this case, the frequency resource which can be allocated to this user is restricted by the BSs with the most heavy load in its cooperative set [6]. In [7], there is no consideration about the resource blocks(RB) constraint where they assume that each user can get an RB from one BS. In the real scenario, however, every BS has different loads causing different RB allocations for users. Therefore, we need to consider the cooperation set and resource allocation together.

In our study, we focus on cooperation set selection and resource allocation in UDN with CoMP. Jain fairness is one of the most important metrics to evaluate the performance of a mobile communication network. To improve the fairness of users, our object is to maximize the minimum sum rate of users, which is named as max-min criteria [8]. Following that, we propose Load-aware Binary Genetic Algorithm(LBGA) to solve the problem of cooperation set selection and Resource Balancing-oriented RB Allocated Algorithm(RB-RBAA) furthermore. System-level simulation is conducted to prove that our algorithm has improved the performance of the minimum user's sum rate.

2 System Model

In this paper, picocell BSs are considered as the BSs in the homogeneous network. Poisson Point Process(PPP) is employed for modeling the distribution of both users and small BSs. The density of users and picocell BSs is λ_{UE} and λ_{SC}, respectively. A user is called as "center user" if it is connecting with only one BS, and "CoMP user" if it is connecting with more than one BSs. In our model, BSs are able to acquire the statistical Channel State Information(CSI) during the process of information exchange.

Assume that the entire set of users in the system is M_Ω, and the set of users associated with BS b is M_b. All the BSs in the system construct set B_Ω, and the user $i(1 \leq i \leq |M_\Omega|)$ has a cooperation set expressed by B_i. For a specific CoMP user i, the SINR can be calculated by

$$\gamma_i = \frac{\sum\limits_{b \in B_i} P_t H_b^i}{\sum\limits_{b' \in B_\Omega \backslash B_i} P_t H_{b'}^i + \sigma^2} \tag{1}$$

In (1), H_i^b represents the channel gain from BS b to user i. $B_\Omega \backslash B_i$ is other BSs without the ones in cooperation set B_i in the system. The transmit power of a picocell BS is P_t. σ^2 is the variance of Additive White Gaussian Noise(AWGN).

[6] implies that the number of BSs in a cooperation set is always no more than 4, since a large number of cooperative BSs will cause serious network throughput loss. In our system, for the user i, the BSs that offer the biggest four RSRPs are denoted by BS_{i1}, BS_{i2}, BS_{i3}, BS_{i4}, and these four BSs compose the candidate cooperation set for the user. In this paper, communication resources refer to the RBs. User i can obtain the proportion of resources β_i from its cooperation set. For a CoMP user, the information will be transmitted on the same RBs, so the proportion β_i will be restricted by the BS which offers the least resource:

$$\beta_i = \min\{\frac{1}{M_b} | b \in B_i\} \tag{2}$$

Thus, the utility function for user i is the sum rate that it can obtain. The function can be calculated by Shannon equation:

$$U_i = \beta_i W \log_2(1 + \gamma_i) \tag{3}$$

where W is the bandwidth of each picocell BS. We use $U_{\min}$ to represent the minimum value of U_i. The utility function of the network throughput is the sum of all the users' utility $U_{total} = \sum_i U_i$.

We also take use of the Jain fairness index to measure the load of the BSs and the sum rate of users in the network [9], which are called load fairness and user fairness accordingly. Load fairness F_B is able to reflect the extent of load balancing.

$$F_B = \frac{\left(\sum_b |M_b|\right)^2}{|B_\Omega| . \sum_b |M_b|^2} \tag{4}$$

User fairness F_U is another index we use to indicate the satisfaction degree of users besides $U_{\min}$ with max-min criteria. It can be expressed as follows:

$$F_U = \frac{\left(\sum_i |U_i|\right)^2}{|M_\Omega| . \sum_i |U_i|^2} \tag{5}$$

3 Problem Formulation and Algorithms

3.1 Mathematical Formulation

To further improve the fairness of users, max-min criterion is considered in this paper. Taking account of restrict frequency resource at each picocell BS, the problem we are considering can be formulated as

$$\begin{aligned} &\max \quad U_{\min} && && (6)\\ &s.t.\ C1: 1 \le |B_i| \le N_{\max} && (i = 1, 2 \cdots |M_\Omega|)\\ &\quad C2: RSRP_{BS_i} \ge RSRP_{BS_{N_{\max}}} && (i = 1, 2 \cdots |M_\Omega|)\\ &\quad C3: 0 \le \beta_i \le 1 && (i = 1, 2 \cdots |M_\Omega|) \end{aligned}$$

The constraint conditions are depicted as follows:

(1) C1 restricts the size of cooperation set for each user. $N_{\max}$ means the largest cooperation set for a CoMP user, which is usually three or four.

(2) C2 implies the candidate cooperation set is restricted in the $N_{\max}$ maximum BSs with biggest RSRPs. If the RSRP is too trivial for a user, it will hardly improve the sum rate for him so we ignore the other BSs.

(3) C3 means that the proportion of the resources that users can obtain from the BS is from 0 to 1.

Since this problem considers discrete variables B_i and continuous variables β_i together which is an NP-hard problem, we cannot figure out the optimal solution directly. Hence, we decompose the problem in two steps. First, on the assumption of average resource allocation, we proposed LBGA to obtain the cooperation set selection results. Second, based on the results in the previous step, RB-RBAA is proposed to further improve $U_{\min}$.

3.2 Load-Aware Binary Genetic Algorithm(LBGA)

Genetic algorithm is an important heuristic algorithm, which makes use of reserving excellent genes in every iteration to find out the optimal solution. Inspired by that, we propose the LBGA to solve the cooperation set selection problem. First of all, we will find out the $N_{\max}$ BSs for every user and create a table mapping user i to its candidate set B_i. In this table, every user has $N_{\max}$ BSs to choose which will be expressed by 0 or 1. For a user i, 0 means the BS is not in its cooperation set and 1 means the opposite. Hence, we take the table which reflects the solution of cooperation set results as a chromosome. The length of a chromosome is equal to the number of users in the network. A gene represents a user. Each gene has $N_{\max}$ units, which could be taken as 1 or 0 representing whether user i connects the corresponding BS or not.

After the modeling of chromosomes, we will use LBGA to iterate until the algorithm converges. The concrete steps of the LBGA are as follows:

(1) For user i in the system, we select the $N_{\max} = 4$ BSs with strongest RSRPs. After mapping cooperation set results to chromosomes, a fixed quantity of chromosomes are generated. In every chromosome, 0 or 1 are randomly generated only if each gene has one "1" on it at least. Notably, for the sake of convenient crossover, the number of chromosomes is even which is expressed as *pop*.

(2) The next operation is to calculate the fitness value $U_{\min}$ which is based on load-aware methods. After that, better chromosomes are selected for heredity and reserving excellent genes. In this step, elitism retention strategy and roulette wheel selection are employed. The chromosomes with best fitness are duplicated to two chromosomes and the others are selected as the order of fitness value. The chromosomes with better fitness value have higher probabilities to be chosen.

(3) Assume that the crossover probability is p_{cross} and arbitrary number of genes may cross with another. On a gene, single point crossover is adopted. It is worth noting that the best two chromosomes will never crossover. Then, the

chromosomes will mutate with the probability of mutation p_{mutate}. Also, we adopt the single-point mutation in our algorithm.

(4) Go back to Step(2) until the fitness value achieves convergency.

After executing the LBGA, we get the cooperation set result for user i with consideration of the load on BSs.

3.3 Resource Balancing-Oriented RB Allocated Algorithm(RB-RBAA)

After obtaining the cooperation set results, we propose RB-RBAA to solve RB allocation problem. Before we deal with the problem, we define some important variables:

Algorithm 1. Resource Balancing-oriented RB Allocated Algorithm(RB-RBAA)

1: Find out all the users' neighbor set N_i
2: Calculate U_i, and then find out user m with $U_{\min}$
3: **for** $k = 1$ to K **do**
4: Find out N_m of user m on $min\_SC(m)$
5: **if** the RB number of the users in N_m is all 1 **then**
6: **goto** 15
7: **end if**
8: **if** there is non-decisive RB for neighbor users in N_m **then**
9: Give an RB to user m from the neighbor user with biggest sum rate
10: **else**
11: **goto** 15
12: **end if**
13: **end for**
14: Calculate U_i, and then find out user m with $U_{\min}$
15: Find out the users on $min\_SC(m)$ with the largest sum rate N_m_first
16: **if** there is only one RB of N_m_first on $min\_SC(m)$ **then**
17: Find out the user with second largest sum rate N_m_second
18: **if** there is only one RB of N_m_second on $min\_SC(m)$ **then**
19: End **Algorithm 1**
20: **end if**
21: **else**
22: Give an RB to user m from N_m_first
23: **end if**
24: Calculate U_i, and then find out user m with $U_{\min}$

(1) Neighbor users set N_i of user i: B_i is the cooperation set for user i, and all the other users on B_i composes the neighbor users set of user i.

(2) Minimum BS for user i: The BSs in B_i which offers fewest RBs is defined as minimum BS for user i, denoted by $min\_SC(i)$.

(3) Decisive resource blocks: The fewest resource blocks that $min\_SC(i)$ can offer are called decisive resource blocks.

(4) Beta Table: It is a table whose size is $|M_\Omega| \times |B_\Omega|$. The number of RBs which user i uses on BS b is stored in the position (i, b) of Beta Table.

After LBGA, since the cooperation set results are determined and the resources are allocated evenly, β_i can be calculated for user i. We can obtain the concrete number of RBs N_{RB} which a user gets. To balance the resources and enhance $U_{\min}$ further, we propose the RB-RBAA to allocate RB for each user. K is the maximum number of iterations. We use index $m(1 \leq m \leq |M_\Omega|)$ to represent the user with $U_{\min}$. We depict the RB-RBAA as **Algorithm 1**.

4 Simulation Results

To simulate an UDN, we choose a topological plane of $1\,\text{km}^2$ with $\lambda_{UE} = 200/\text{km}^2$ and $\lambda_{BS} = 40/\text{km}^2$. The number of chromosomes *pop* is 300 with $p_{cross} = 0.5$ and $p_{mutate} = 0.08$. $W = 10\,\text{MHz}$ which means that a BS has 50 RBs. We employ the WINNER model to generate H.

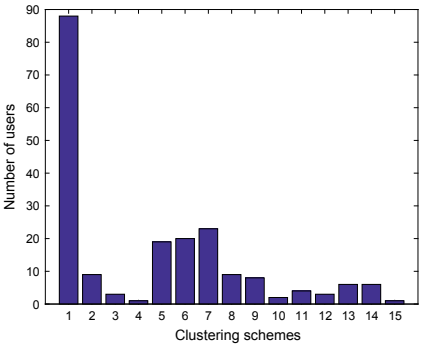


Fig. 1. Users distribution

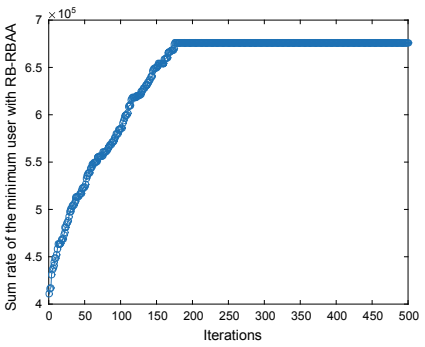


Fig. 2. $U_{\min}$ with RB-RBAA

Figure 1 shows the cooperation set selection schemes that each user chooses. As every user selects $N_{\max}$ BSs for its candidate BSs, there are 15 kinds of cooperation set selection schemes for each user. As Fig. 1 shows, "1" represents that the users who select their BS_1 as their serving BSs, "2" represents that the users who select their BS_2 as their serving BSs and so on. Similarly, "15" means that the users choosing BS_1, BS_2, BS_3, and BS_4 as their cooperation set finally. Compared with cooperation set selection schemes with a fixed number of cooperation BSs, this method will be more flexible. In this way, the cooperation set selection scheme is easy to adapt for the change of topology and channel conditions.

Figure 2 shows that $U_{\min}$ is enhanced further by RB-RBAA with a small number of iterations. As the cooperation set has been confirmed after LBGA,

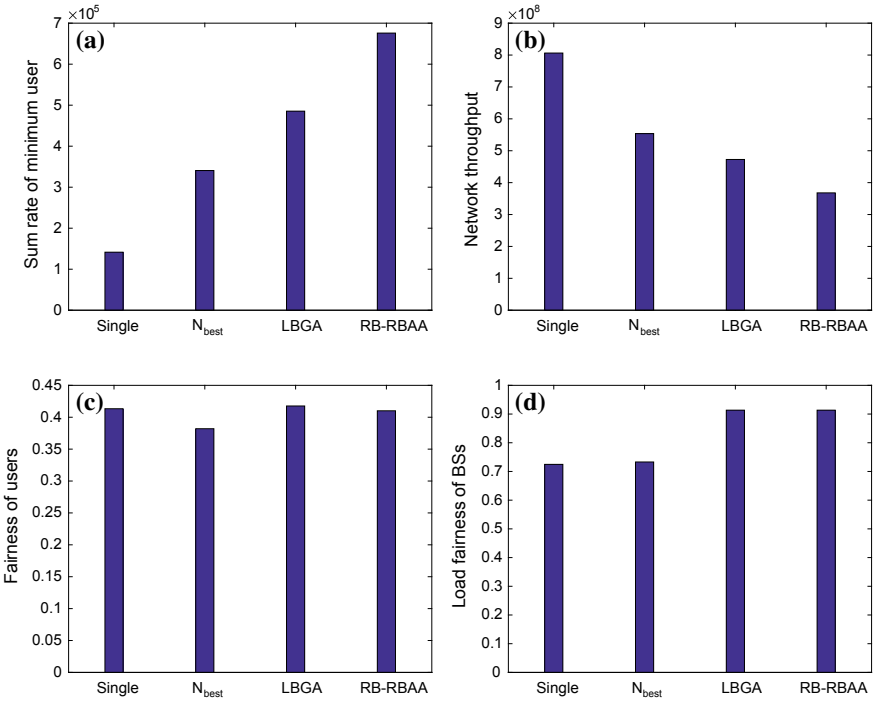


Fig. 3. Performance comparison within four algorithms

the load fairness remains unchanged. It is worth noting that we only consider the proportion of resource blocks that a user will acquire with no consideration of specific RB allocation.

As shown in Fig. 3, four metrics $U_{\min}$, F_U, F_B and U_{total} are compared among four different algorithms. In N_{best} scheme, a CoMP user will choose the fixed number BSs as its cooperation set. We simulate the performance of N_{best} scheme with $N = 3$ with 30% CoMP users in the network for comparison. The single scheme in the paper means that a user will just connect with one BS for communication. Simulation results have shown that our algorithms perform better than N_{best} scheme on $U_{\min}$ in Fig. 3(a). Through LBGA, the $U_{\min}$ improves 42.5% compared to N_{best} scheme. The metric further improves 39.2% after RB-RBAA compared to LBGA. Meanwhile, F_B is ameliorated in Fig. 3(d). Compared with N_{best} scheme, F_U with Jain indicator also increases in Fig. 3(c). We can find out that F_U changes little between our two algorithms and single scheme which means our algorithms do not lose Jain fairness as well. However, comparing with single scheme and N_{best} scheme, our objective improves at the cost of the decreased network throughput U_{total} in Fig. 3(b).

5 Conclusion

Fairness is an important metric to measure the performance of a system. To improve the fairness, LBGA and RB-RBAA are proposed in this paper to solve the cooperation set selection with resource allocation. Simulation results show that our algorithms perform better compared with traditional single scheme and N_{best} scheme. LBGA can make the cooperation set selection more flexible for users, which also benefits for ameliorating the load fairness of BSs. Moreover, RB-RBAA improves $U_{\min}$ efficiently and guarantees the fairness of users.

Acknowledgments. This work has been supported in part by the National Natural Sciences Foundation of China(NSFC) under Grants 61501140, 61701136 and 61525103.

References

1. Bhushan, N., Li, J., Malladi, D., Gilmore, R.: Network densification: the dominant theme for wireless evolution into 5g. Commun. Mag. IEEE **52**(2), 82–89 (2014)
2. Kamel, M., Hamouda, W., Youssef, A.: Ultra-dense networks: a survey. IEEE Commun. Surv. Tutor. pp(99), 1–1 (2017)
3. E. U. T. R. Access: Further advancements for e-utra physical layer aspects. 3GPP Techn. Specifi. TR, **36**, V2 (2010)
4. Chen, S., Zhao, T., Chen, H.H., Lu, Z., Meng, W.: Performance analysis of downlink coordinated multipoint joint transmission in ultra-dense networks. IEEE Netw. pp(99), 12–20 (2017)
5. Tanbourgi, R., Singh, S., Andrews, J.G., Jondral, F.K.: A tractable model for noncoherent joint-transmission base station cooperation. Wirel. Commun. IEEE Trans. **13**(9), 4959–4973 (2014)
6. Liu, L., Garcia, V., Tian, L., Pan, Z., Shi, J.: Joint clustering and inter-cell resource allocation for comp in ultra dense cellular networks. In: IEEE International Conference on Communications, pp. 2560–2564 (2015)
7. Garcia, V., Zhou, Y., Shi, J.: Coordinated multipoint transmission in dense cellular networks with user-centric adaptive clustering. Wirel. Commun. IEEE Trans. **13**(8), 4297–4308 (2014)
8. Bertsekas, D., Gallager, R.: Data Networks, 2nd ed. Prentice-Hall, Inc. (1992)
9. Jain, R., Chiu, D., Hawe, W.: A quantitative measure of fairness and discrimination for resource allocation in shared computer systems,. Comput. Sci. vol. cs.ni/9809099 (1998)

An Anomaly Detection Model Based on Immune Network with Variable Any-R-Intervals Matching Rule

Xiaoke Feng(✉), Zhuo Sun(✉), and Bin Li(✉)

Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing, China
{xiaokefeng,binli}@bupt.edu.cn, zhuosun@163.com

Abstract. When the immune network is used in anomalistic electromagnetism signals detection, a critical issue emerged that lots of "detection holes (gaps of the detector coverage compare to the target area)" are caused by fixed-size detectors and invariable matching threshold. This paper improves the original model by increasing the detectors' coverage and proposes an anomalistic electromagnetism signal detection model based on an immune network with variable any-r-intervals matching rule. The sizes and matching thresholds of different detectors will learn from the training set in this model, which can reduce the detection holes obviously. The model evaluated by wireless signals and the experiment results show that the model can reduce detection holes and improve detection accuracy compared with the original model.

Keywords: Anomaly detection · Immune network · Variable · Any-r-intervals

1 Introduction

Wireless communication technology is widely used in civil and military communication. To protect the security of electromagnetism space, the key is to find anomalistic signals from complex electromagnetic space composed of legal signals, random noise, interfering signals and so on. While the main challenge is that different parameters of different signals have different fluctuation range in the actual environment due to the weather, geographical conditions and so on.

The authors of [1] first introduce AIS (artificial immune system) into anomalistic electromagnetism signal detection. The results in [2] apply immune network to solve the problem that the amount of the detectors will be very large if the self-set space is too small. Research in [3] makes use of IDT (immune danger theory) to reduce the random noise's impact on the performance of detection model. However, the models mentioned in the above researches [4, 5] commonly use a matching rule based on Euclidian distance with fixed-sized detectors and invariant matching threshold, which lead to a large number of detection holes. That is the main factor affecting the performance of the detection models. Considering that V-detector algorithm and variable *rcb* matching rule [6, 7] can reduce the amount of detection holes, a model using variable any-r-intervals matching rule is proposed in this paper to solve the aforementioned problem.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 367–375, 2019.
https://doi.org/10.1007/978-981-13-6264-4_45

The rest of this paper is organized as follows: Sect. 2 describes the problem in previous studies and puts forward a corresponding solution. Section 3 demonstrates the feasibility of the method proposed in this paper. In Sect. 4, the whole anomaly detection process is introduced in detail. In Sect. 5, experimental results are presented and discussed. Conclusion and future work are drawn in Sect. 6.

2 Problem Formulation

The electromagnetism signal detection model based on the immune network [2] focus on generating a set of mature detectors with the same coverage radius through cloning and selection algorithm and inhibition. The main concerning of this approach is that the false alarm rate is low but relatively the false negative rate is high. In the complicated electromagnetic environment, the detection model should focus on that the fluctuation ranges of different signals are different. Although in the training process, detectors will learn from the self-set adaptively, in the detection module the threshold is still invariable.

The matching between the antibody and antigen is uncertain and fuzzy due to the diversity of the immune system [6]. If the size of the mature detector is too big, false negative rate will be high, and if too small, false alarm rate will be high. Similarly, if the matching threshold is too high, the coverage rate will decrease, and if the matching threshold is too low, the self-space will be over-covered. Therefore, this paper focus on using the variable any-r-intervals matching rule instead of the matching rule based on Euclidian distance. A mature detector described with two parts, a matrix $M_{2\times n}$ and an integer r. Our goal is to find the most suitable $M_{2\times n}$ and r for every semi-mature detector.

3 Solution Based on Variable Matching Rule

Next, we theoretically analysis why variable any-r-intervals matching rule can reduce detection holes and improve the performance of the detection model in two parts.

3.1 The Matrix $M_{2\times n}$

The value of matrix $\boldsymbol{M_{2\times n}}$ determines its ability to cover the self-set space. The shape of a mature detector is a hyper-rectangle. Theoretically, there is also a $M_{2\times n}$ can make the model get the best performance at the condition that the value of $M_{2\times n}$ is fixed. Under these circumstances, the method to improve the performance of detection model is to reduce the detection holes which caused by fixed matrix $M_{2\times n}$. This paper will develop further training for the semi-detectors with the training set and get the optimal matrix threshold $M_{2\times n}$ of every semi-detector.

We execute simulation experiments based on fixed $M_{2\times n}$ and variable $M_{2\times n}$ with the same training set. In Figs. 1 and 2, a small black rectangle represents a detector and its cover space, the area in the red line is self-set space. The number of detectors in Fig. 2 is the same as Fig. 1. It is obvious that the detection holes in Fig. 2 is much less than Fig. 1, specifically, detectors with suitable size matrix M can cover the self-set space better than detectors with same size matrix M.

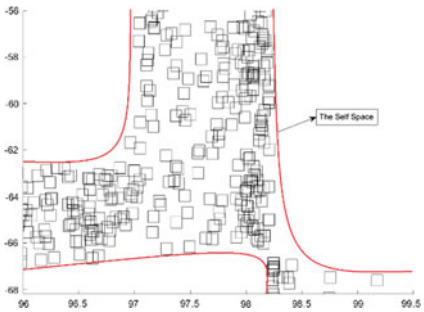


Fig. 1. Detectors with same size matrix M

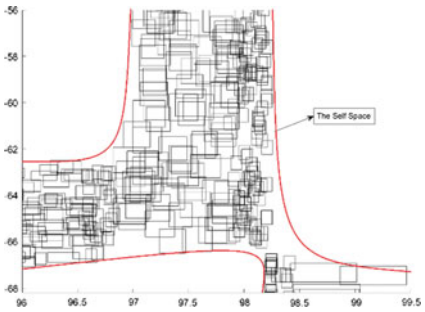


Fig. 2. Detectors with suitable size matrix M

3.2 r-Matching Threshold

In the process of Clone Selection Algorithm, the parameters' matching number of every pair of best-matched self-signal and detector are different [8]. If a same matching threshold r is set for all detectors, it is hard to decrease both the false alarm rate and false negative rate. In research [9], Esponda gives the calculation method of covering space of *rcb* matching rule. The calculation method of covering spaces of any-r-interval matching method is proposed in this paper as follows.

A detector w contains L parameters, to compute detector w' coverage, giving following definitions:U is the complete set of strings of L length, every parameter may have N values because of real number coding method. r is the matching threshold of any-r-interval matching rule, and $r > 1$.K(L) stands for the set of strings can match with w based on the matching threshold r. The K(L) can be got through permutation and combination formula.

$$\mathrm{K(L)} = A_L^r \cdot N^{L-r} = \frac{L!}{(L-r)!} \cdot N^{L-r} \tag{1}$$

U, the complete set of strings of L length subtracts the K(L) is the set of strings can't match with w base on the matching threshold r, we define it as $\mathrm{C}(L)$. It is obvious that U can be calculated as Eq. 2:

$$\mathrm{U} = N^L \tag{2}$$

So that $\mathrm{C}(L)$ can be obtained by Eq. 3:

$$\mathrm{C}(L) = U - K(L) = N^L - \frac{L!}{(L-r)!} \cdot N^{L-r} \tag{3}$$

Take the derivative of K(L) and get its monotonicity: when $r > L - N + 1$, K(L) is monotonically decreasing with the increase of r. The monotonicity of C(L) is contrary to K(L).

Because N can be infinite in real-value coding, we suppose that it is big enough to guarantee that $r > L - N + 1$, so C(L) is monotonically increasing. If the value of r is smaller, the detection holes will decrease. For the semi-mature detectors generated from clone selection algorithm, the matching number can be smaller than r, and the covering space of these detectors will be bigger.

4 Anomaly Detection Model Based on Any-R-Intervals Matching Rule

As shown in Fig. 3, the detection model proposed in this paper contains three main parts: training module, detection module, and adaptive update module. Generally speaking, the ideal set of background signals contains no anomalistic signals can be used to training the detection network. The most significant improvement in this paper is the training module. Different from the research in [2], the detectors generated by clone selection and inhibition are regard as semi-mature detectors instead of mature detectors, there is a second training and then the semi-mature detectors become mature detectors. The detection module consists of mature detectors and adaptive update module.

4.1 Parameters Setup

Electromagnetism signal has many parameters such as frequency, coordinates of station, modulation and so on. As for the model in [2], signal's parameters are binary-encoded. It has many advantages, for example, many kinds of data can be expressed in binary form. But the limitation is that it is hard to explain the potential scalability [6, 10]. So real-value coding is necessary for artificial immune system.

Matching rule is the foundation of all classification, recognition and detection algorithm. Research [2] uses Hamming distance based on binary form, it chooses Euclidian distance for real number coding. Euclidian distance is defined as follows:

Pending Signals: $\{P = (x_1, x_2, \ldots, x_n)\}$

Detectors: $\{Q = (y_1, y_2, \ldots, y_n)\}$

$$\text{Distance} = \sqrt{\frac{\sum_i^n (x_i - y_i)^2}{n}} \tag{4}$$

The signals in the complicated electromagnetic environment are mutable. There will be errors when receiving signals and the misalignment exists in actual measurement. Besides, the parameter of a normal signal is fluctuant within a certain interval, the chosen intervals constitute a hyper-rectangle. But the matching rule based on Euclidian distance covers a hyper-sphere. Obviously that using a hyper-sphere to cover a hyper-rectangle is not a proper choice. In this paper, the any-r-intervals matching rule uses

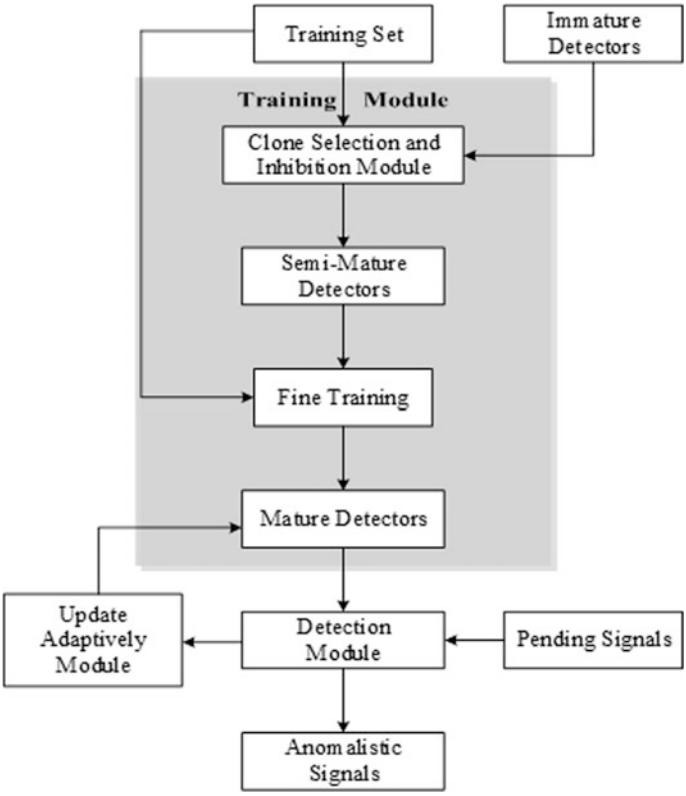


Fig. 3. Improved anomaly detection model

real-value coding instead of binary coding. The detailed description of any-r-intervals matching rule is given by Table 1. The input and preset threshold are shown as follows:

Pending Signals: $\{\mathrm{P}|\mathrm{P}_l = (x_{l,1}, x_{l,2}, \ldots, x_{l,n})\}$

Mature Detectors: $\{M_{2\times n}, \mathrm{r}\}$.

Matching Threshold: R

4.2 Training Process

In the training step, we get the semi-mature detectors using clone selection algorithm [2]. The semi-mature detectors can be expressed as: $\{\ \mathrm{D}|D_h = (d_{h,1}, d_{h,2}, \ldots, d_{h,n}), h = 1, 2, \ldots, \mathrm{N}_1\}$. Then train the semi-mature detectors to get mature detectors which have their own size matrix and matching threshold.

In the process of fine training, we match a training signal with every semi-detector and find the semi-detector which is matching the training signal best, then get the value of affinity r', if r' is smaller than the present r of the semi-detector, replace r with r'. Next, we set the interval of each parameter of the semi-detectors. If the parameter of the training signal is beyond the initial interval of the semi-detector, replace the interval's

Table 1. Pseudo Code of Any-r-intervals matching rule

Input: Pending Signals and Mature Detectors

for: P = 1 to num_1

for: M = 1 to num_1

for: $parameters$ = 1 to n

for pending signalP_l , calculate how many parameters fall into the corresponding intervals of every detector

if: $max\{count(l,:)\} \geq R, P_l \in normal\ signals$

else: $P_l \in anomalistic\ signals$

Output: Label of every pending signal.

upper or lower bound with the corresponding value of the training signal. The algorithm of fine training is described in Table 2.

Table 2. Pseudo code of fine training

Input: Training signals; Semi-mature detectors

for S: $t = 1$ to N_2

for: D: $h = 1$ to N_1

for:$parameters$: $i = 1\ to\ n$

Calculate how many parameters fall into the corresponding intervals of every semi-mature detector

if: $max\{count(l,:)\} \leq R$, use $max\{count(l,:)\}$ instead of D_h's R and use the parameters fall out of the corresponding intervals of S_t update D_h's size matrix

end

Output: Mature detectors with different size matrix and r.

The input of fine training and their format are described as follows: Training Signals: $\{ S|S_t = (s_{t,1}, s_{t,2}, \ldots, s_{t,n}), t = 1, 2, \ldots, N_2\}$. An initial size matrix $M' = \begin{vmatrix} a_1 & \ldots & a_n \\ b_1 & \ldots & b_n \end{vmatrix}$ and initial threshold $\{R = [r, r, \ldots, r]_{1\times n}\}$. Based on the initial size

matrix and semi-mature detectors we can get the initial intervals: $\left\{ M|M^h = \begin{vmatrix} d_{h,1} - a_1 & \dots & d_{h,n} - a_n \\ d_{h,1} + b_1 & \dots & d_{h,n} + b_n \end{vmatrix} \right\}$.

4.3 Detection

We will detect pending signals with mature detectors and update the set of mature detectors according to the detection results. If the pending signals can't match with any detector, report it as anomalistic signal. If all parameters of a pending signal fall into the corresponding intervals of a mature detector, update the mature detector set with the pending signal.

5 Experimental Results

Data detected from several radios are used to verify the performance of any-r-intervals matching rule. The signals that received from radios in the real electromagnetic space are regarded as background signals. The signals that received from our own signal source, which emitter signals with different frequencies, bandwidth, directions, and other parameters, are regarded as anomalistic signals. The experiment data is divided into training data only contains normal signals and pending data which contains both normal signals and anomalistic signals. The details can be got in Table 3.

Table 3. Experiment data

	Normal signals	Anomalistic signals
Training data	50	0
Pending data	100	100

This paper evaluates the performance of the detection model using different matching rules in terms of three standards: FNR、FAR and Accuracy. *FN* is the number of anomalistic signals that detected to be normal signals, *TP* is the number of anomalistic signals that detected to be anomalistic signals. *TN* is the number of normal signals that detected to be normal signals, *FP* is the number of normal signals that detected to be anomalistic signals.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \tag{5}$$

$$\text{FNR(False Negative Rate)} = \frac{FN}{FN + TP} \tag{6}$$

$$\text{FAR(False Alarm Rate)} = \frac{FP}{FP + TN} \tag{7}$$

From Fig. 4, we can see that the FNR has been decreased by about 1.5% points and FAR has been decreased by 12.4% points. The total accuracy has been increased by 7.65% points. The performance of any-r-intervals matching rule is improved compared with the model based on Euclidian Distance. It also proves that the hyper-rectangle detectors' ability to cover the hyper-rectangle space is better than hyper-sphere detectors. And any-r-intervals matching rule's tolerance to the large fluctuation of certain parameter caused by objective factors is necessary.

From Fig. 5, it is obvious that the FNP and FAR of the detection model proposed in this paper are decreased compared with the model based on the immune network. And the accuracy is also improved. The detection model proposed in this paper solves the problem of detection holes caused by the immune network to a certain degree.

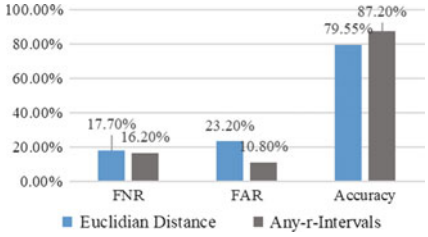


Fig. 4. Emulation result of the two matching rules

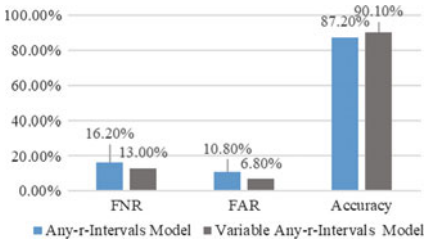


Fig. 5. Emulation result of the two models

6 Conclusion

AIS model used in anomaly detection has a long period of development, but problems still exist in introducing the model into electromagnetic spectrum monitoring. This paper proposes a method based on variable any-r-interval matching rule to solve the aforementioned problem and improves the performance through reducing the detection holes. The experiment with monitoring data of radio signals is performed to verify the validity of our model. Besides, on the condition that detection performance will not change, whether the number of detectors can be reduced, will be discussed in the future work.

References

1. Jie, M., Ying-chun, S., Zi-fa, Z., et al.: An anomalistic electromagnetism signal detection model based on artificial immune system. In: 2010 International Conference on Communications and Intelligence Information Security, pp. 256–260 (2010)
2. Jie, M., Shi, Y.-C., Zhong, Z.-F., et al.: An anomalistic electromagnetism signal detection model based on immune network. Appl. Res. Comput. **28**(6), 2140–2143 (2011)

3. Chuanjin, Lin, Jie, Ma., Xianghui, Yuan: An anomalistic electromagnetism signal detection algorithm in noisy environment based on immune danger theory. Ship Electron. Eng. **11**, 60–64 (2011)
4. Hou, H.Y., Dozier, G.: An evaluation of negative selection algorithm with constraint-based detector. In: ACM Southeast Regional Conference 2006, pp. 134–139. Melbourne, Florida, USA (2006)
5. Research on the generation algorithm of intrusion detector based on artificial immunity. Chongqing, Chongqing University (2008)
6. Hong-wei, M.O., Xing-quan, Z.U.O.: Artificial Immune System. The Science Publishing Company, Beijing (2009)
7. Ji, Z., Dasgupta, D.: Real-valued negative selection algorithm with variable-sized detectors. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2004), pp. 287–298. ACM, Seattle, WA, USA (2004)
8. Ji, Z., Dasgupta, D.: An efficient negative selection algorithm with "probably adequate" detector coverage. Inf. Sci. **179**, 1390–1406 (2009)
9. Esponda, F., Forrest, S.: Detector coverage under the r-contiguous bits matching rule. Technical Report TR-CS-2002–03, University of New Mexico (2002)
10. Pimentel, M.A., Clifton, D.A., Tarassenko, L.: A review of novelty detection. Sig. Process. **99**, 215–249 (2014)

Multiparameter Analysis for Distributed Storage Schemes in Wireless D2D Networks

Juan Li, Shushi Gu(✉), Ye Wang, Yue Li, and Qinyu Zhang

Communication Engineering Research Center, Harbin Institute of Technology, Shenzhen, Guangdong, China
{lijuan2016,liyue}@stu.hit.edu.cn,
{gushushi,wangye83,zqy}@hit.edu.cn

Abstract. In this paper, we consider a novel wireless network for content distribute storage, where devices could use device-to-device (D2D) communication to transmit data. Several storage schemes, such as maximum distance separable (MDS) code and regenerating code, are employed for content downloading and repairing as node left and data lost. On the basis of multiparameter analysis, we found that targeting on the popular file caching, smaller download locality is the most suitable to the system based on fixed storage nodes and lower repair frequency. Moreover, MBR codes do not apply to D2D distributed storage due to high download cost, and MDS code is a convenient and appropriate scheme for majority practical D2D networks.

Keywords: D2D communication · MDS codes · Repair bandwidth · Repair interval

1 Introduction

Rapidly increasing global data causes a heavy challenge for overwhelmed wireless communication in 5G networks. Fortunately, D2D communication is an attractive technology lately proposed to reduce base station (BS) load and increase throughput in cellular cell [1]. In the recent years, some literature study that content can be distributed and cached in modern smart mobile devices and users download it from the storage nodes using D2D communication [2–4]. When data is cached on mobile devices with an encoded manner, we can easily contact it with a distributed storage system (DSS) by D2D links. However, the problem of repairing lost data must be considered, as high-speed mobility and frequent switching of the crowd and the vehicles.

The combining of D2D and DSS has been studied in [5,6]. Communication cost caused by data downloading and repairing was analyzed in [5]. Lost data is instantaneous repair, and user can always download content from storage node using D2D communication. 2-replication can get minimal communication

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 376–384, 2019.
https://doi.org/10.1007/978-981-13-6264-4_46

cost when lost data is repaired instantaneously. While [6] analyzes the relationship between communication cost and repair interval in a wireless network scenario, they gave the impact of request rate and the number of storage nodes (repair locality) connected to repair on total communication cost, but they do not analyze the impact of the number of nodes and the number of storage nodes (download locality) connected during download on communication cost.

In this paper, we give the analysis for multiparameter of several schemes in D2D content storage networks, as the number of nodes and download locality impacting on total communication cost, i.e., replication, MDS codes, and regenerating codes. We find that distributed storage can achieve communication cost and we show when files are frequently downloaded, communication cost using distributed storage is always lower than that from BS, and low download locality can get lower communication cost.

2 System Model

We consider a wireless network as shown in Fig. 1 served by a stationary BS and mobile devices entering and leaving network according to a Poisson random process. Nodes in network can store data and use D2D communication for content delivery. When storage nodes cannot complete download or repair process using D2D communication, it has to be performed by BS. In Fig. 1, blue nodes are storage nodes, green node is the requester. Gray nodes represent empty nodes and red node represents a new helper node to be repaired by others.

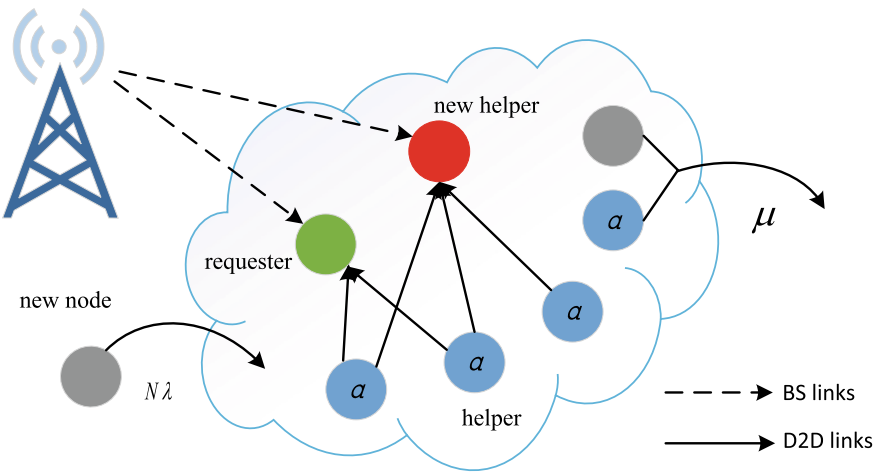


Fig. 1. A wireless network system model

The initial number of nodes is N. For simplicity, we assume that there is only one file with size F in BS. The file is divided into k blocks and encoded into

n coded symbols, using (n, k) erasure coded with code-rate of $R = k/n$, stored in m storage nodes, and $m \leq n$. When a storage node departs the network, it needs to connect to at least r storage nodes to repair the lost data, and h storage nodes need to be connected when downloading file. Moreover, we assume that each storage node equally stores α bits of data.

The expected departure rate of a node is μ. T_s is defined as a node duration time, which is an independent, identically, distributed (i.i.d.) exponential distribution with pdf

$$f_{Ts}(t) = \mu e^{-\mu t}, t \geq 0 \tag{1}$$

Nodes in the system randomly request files at a rate ω, and the request interval time T_r is i.i.d. exponential distribution with pdf

$$f_{Tr}(t) = \omega e^{-\omega t}, \omega \geq 0, t \geq 0 \tag{2}$$

When storage node number $i \geq h$, the user downloads content transmitting $h\alpha \geq F$ bits using D2D download, and retrieving F bits from BS otherwise. For simplicity, we assume that download bandwidth of all nodes is the same. ρ_{BS} and ρ_{D2D} are defined as the cost units of transmitting 1 bit data (c.u./bit) from BS and from storage node, respectively, and $\rho = \rho_{BS}/\rho_{D2D} > 0$ is defined as cost ratio. The meaning of $\rho > 1$ is that traffic load in BS-link is much higher than D2D-links. γ is defined as the amount of data when repairing lost data which is referred to as repair bandwidth, and the data is from r storage nodes, and each node provides $\beta \leq \alpha$ bits. γ_{BS} and γ_{D2D} are defined as repairing one failed node by BS and storage, nodes, respectively, and $\gamma_{BS} = \alpha$.

3 Distributed Storage Schemes

Erasure codes for DSS usually are described by the number of nodes m, download locality h, and repair locality r, i.e., $[m, h, r]$. In this section, we mainly consider n-replication, MDS codes, and regenerating codes for D2D storage network.

3.1 n-Replication and MDS Codes

For erasure codes, each storage node stores one coded symbol, i.e., $m = n$, $\alpha = F/k$. Due to MDS property, the number of storage nodes that need to be connected for D2D download and repair is $h = r = k$. In addition, $\beta = \alpha = F/k$ and $\gamma = F$. The simplest MDS code is n-replication, each storage node stores the full file, i.e., $\alpha = F$, $r = h = k = 1$.

3.2 Regenerating Codes

For an exact-repair minimum storage regenerating (MSR) code, since one node stores more than one coded symbol, the relationships between the $[m, h, r]$ and (n, k) of MSR code is accurately repaired are $k = h(r-1)$ and $n = m(r-h-1)$,

$r \in [2(h-1), m-1]$ [7], we can get $h \leq (m+1)/2$, so maximum code-rate for MSR is $(1/2 + 1/m)$. We can obtain

$$(\alpha_{MSR}, \gamma_{MSR}) = (\frac{F}{h}, \frac{Fr}{h(r-h+1)}) \tag{3}$$

Minimum bandwidth regenerating (MBR) code reduces repair bandwidth by increasing α, the relationship between the $[m,\ h,\ r]$ and $(n,\ k)$ parameters of MBR code are $k = hr - \binom{h}{2}$ and $n = mr$, $r \in [h, m-1]$, and we can get

$$(\alpha_{MBR}, \gamma_{MBR}) = (\frac{2Fr}{h(2r-h+1)}, \frac{2Fr}{h(2r-h+1)}) \tag{4}$$

4 Communication Cost Analysis

4.1 Download Cost and Repair Cost

In this section, we mainly analyze the relationship between multiparameter and communication cost. $\bar{C}_r$ is defined as the average repair cost, $\bar{C}_d$ is the average download cost, and the average total communication cost is $\bar{C} = \bar{C}_r + \bar{C}_d$. Δ is defined the interval time between two repair processes, $\Delta = 0$ corresponds to instantaneous repair. The number of storage nodes m and the probability p that a storage node is available are Bernoulli distribution, so the probability mass function (pmf) of m and p is

$$b_i(m,p) \triangleq \binom{m}{i} p^i (1-p)^{m-i}, 0 \leq i \leq m \tag{5}$$

Repair Cost. In a repair interval Δ, the probability that the storage node does not leave the system is p, that is, the average lifetime of a node is greater than Δ, and the probability is

$$p = \mathrm{P}(T_s > \Delta) = e^{-\mu\Delta} \tag{6}$$

The probability that there are i storage nodes is $b_i(m,p)$. At this time, $m - i$ nodes need to be repaired. If $i \geq r$, lost data can be repaired using D2D communication, otherwise repaired by BS. The average number of nodes that are repaired by D2D and BS, respectively, are

$$m_r^{D2D} = \sum_{i=r}^{m} (m-i) b_i(m,p), m_r^{BS} = \sum_{i=0}^{m-1} (m-i) b_i(m,p) \tag{7}$$

So, total average repair cost is

$$\bar{C}_r = \frac{1}{\Delta}(\rho_{BS}\gamma_{BS} m_r^{BS} + \rho_{D2D}\gamma_{D2D} m_r^{D2D}) \tag{8}$$

Download Cost. Similar to repair cost, the download cost also has two parts. Define p_{BS} and p_{D2D} as the probability of downloading from BS and from D2D, respectively, $p_{BS} + p_{D2D} = 1$, and define $\mu_i = i\mu, i \in [h, m]$, $p_i = e^{-\mu_i \Delta}$.

In a repair interval, after one node departed, no repair process is performed. Therefore, the number of storage nodes in system can be described by Poisson birth–death process. The probability of there are at least i storage nodes when a request occurred [6] is

$$\mathrm{s}_i = \frac{1}{\Delta} \sum_{i'=i}^{m} \frac{1 - p_{i'}}{\mu_{i'}} \prod_{\substack{j=i \\ j \neq i'}}^{m} \frac{j}{j - i'} \tag{9}$$

And the expected download cost is

$$\bar{C}_d = N\omega(\rho_{BS} F p_{BS} + \rho_{D2D} h \alpha s_h) \tag{10}$$

$\rho_{BS} F$ and $\rho_{D2D} h\alpha$ represent the cost of downloading content once from BS and storage nodes, respectively, $N\omega$ represents the total request rate (in unit time). Combining (Eq. 8) and (Eq. 10), we get the average communication cost as $\bar{C} = \bar{C}_r + \bar{C}_d$.

$\Delta \to 0$ corresponds to instantaneous repair, communication cost is

$$\bar{C} = (N\omega h\alpha + m\mu\gamma_{D2D})\rho_{D2D} \tag{11}$$

4.2 Hybrid Download and Repair

For both download and repair, when download or repair cannot be completed from storage nodes using D2D communication, a node can retrieve part of the data from available storage nodes and the rest of the data is provided by BS. The storage node and the BS cooperate to complete the repair or download process called as hybrid download and repair.

Hybrid Repair Cost. When storage node number $i < r$, repair process cannot be completed only from the storage nodes. However, $i\beta$ bits can be retrieved from the i storage nodes, and the remaining $\gamma_{D2D} - i\beta = (r - i)\beta$ bits are from BS, corresponding repair communication cost of one failed storage node is $(i\beta\rho_{D2D} + (r - i)\beta\rho_{BS})$. For conventional scheme, repair cost corresponds to BS repair with cost $\rho_{BS}\gamma_{BS}$. Hybrid repair can reduce communication cost when $(i\beta\rho_{D2D} + (r-i)\beta\rho_{BS}) < \rho_{BS}\gamma_{BS}$, i.e., $i > \frac{\rho_{BS}}{\rho_{BS}-\rho_{D2D}}(r - \frac{\gamma_{BS}}{\beta})$, we define $r_h = \min\left\{\lceil \frac{\rho_{BS}}{\rho_{BS}-\rho_{D2D}}(r - \frac{\gamma_{BS}}{\beta})\rceil, r\right\}$. For $i \in [r, m-1]$, D2D repair is performed, and when $i \in [r_h, r-1]$, the hybrid scheme is performed, otherwise BS repair is performed. The repair cost is

$$\bar{C}_r^{hybrid} = \frac{1}{\Delta}(\rho_{BS}\gamma_{BS}\sum_{i=0}^{r_h-1}(m-i)b_i(m,p) + \sum_{i=r_h}^{r-1}(\rho_{BS}(r-i)+i\rho_{D2D})(m-i)\beta b_i(m,p) + \rho_{D2D}\gamma_{D2D}\sum_{i=r}^{m-1}(m-i)\beta b_i(m,p)) \quad (12)$$

Hybrid Download Cost. Same as repair, hybrid download communication cost is $(i\alpha\rho_{D2D} + (h-i)\alpha\rho_{BS})$, and the condition of hybrid download reduced cost of $(i\alpha\rho_{D2D} + (h-i)\alpha\rho_{BS}) < \rho_{BS}F$, i.e., $i > \frac{\rho_{BS}}{\rho_{BS}-\rho_{D2D}}(h - \frac{F}{\alpha})$. We define $h_h = \min\left\{\lceil\frac{\rho_{BS}}{\rho_{BS}-\rho_{D2D}}(h-\frac{F}{\alpha})\rceil, h\right\}$. For $i \in [h, m]$, D2D download is performed, and the hybrid download is performed when $i \in [h_h, h-1]$, otherwise, BS repair is performed. From (9), we can know probility of a request arrived is $(s_i - s_{i+1})$ when there are i storage nodes in system. So, the download cost is

$$\bar{C}_d^{hybrid} = N\omega(\rho_{BS}F(1-s_1) + \rho_{BS}F\sum_{i=1}^{h_h-1}(s_i - s_{i+1}) + \sum_{i=h_h}^{h-1}(\rho_{BS}(h-i)+i\rho_{D2D})\alpha(s_i - s_{i+1}) + \rho_{D2D}h\alpha s_h) \quad (13)$$

5 Simulation and Results

We evaluate the overall communication cost of erasure code schemes discussed previously. The parameters of simulated network are set as follows. The leaving rate $\mu = 1$, $\rho_{D2D} = 1$, so cost ratio $\rho = \rho_{BS}$. Normalizing $\bar{C}$ to the cost of downloading from BS, $N\omega\rho$, i.e., $\bar{C}/FN\omega\rho$, when the value of $\bar{C}/FN\omega\rho$ is more than 1, it means DS outperforms downloading from BS.

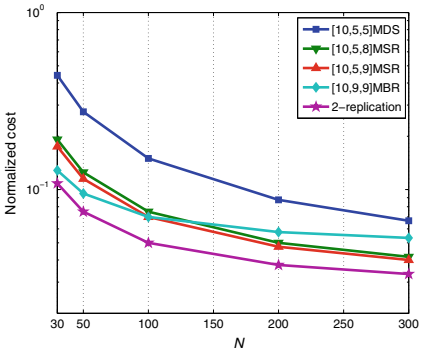


Fig. 2. Normalized cost for instantaneous repair.

We set code-rate $R = 1/2$, per node request rate $\omega = 0.02$, and cost ratio $\rho = 40$. As shown in Fig. 2, it is obvious that the value of $\bar{C}/FN\omega\rho$ is the smallest

with 2-replication scheme, and the value of $\bar{C}/FN\omega\rho$ decreases as user number N increases. This is because the value of $\bar{C}_d/FN\omega\rho$ is a constant, and the value of $\bar{C}_r/FN\omega\rho$ has a negative relationship to N. Particularly, when $N > 150$, MSR code outperforms MBR code. The reason is that downloading a file needs more than F bits for MBR, while other schemes is F bits.

We chose the same parameters as Fig. 2, and set node number $N = 200$. The values of $\bar{C}/FN\omega\rho$ of all erasure code scheme are smaller than 1 in Fig. 3, which shows that caching files with high download frequency in mobile devices can reduce communication cost. The cost is always the highest for MBR due to the high cost of downloading content. When the repair interval is larger, the cost of replication scheme is smallest, that is because the data can be downloaded using D2D communication as long as there is one storage node in the system, and users have a great possibility to download content using D2D communication. It is worth noting that MDS and MSR have equivalent cost for larger repair interval because of the same probability to using D2D downloading content.

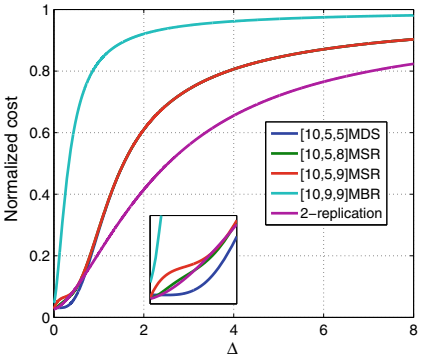


Fig. 3. Normalized cost various Δ

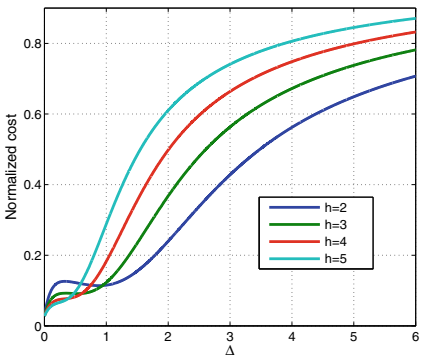


Fig. 4. Normalized cost for [10, h, 9]MSR

Since α has a significant impact on $\bar{C}_d$ and $\alpha = F/h$, we analyze the impact of h on $\bar{C}$ next. As shown in Fig. 4, the value of normalized cost for $h = 5$ is the lowest when Δ is smaller, and the lowest value is achieved for $h = 2$ when Δ is larger. The reason is that larger h results in smaller repair bandwidth γ when fixing storage nodes m and repair locality r. However, the probability of downloading using D2D communication p_{D2D} (Eq. 9) is affected by h, which decreases with increasing h for fixed storage nodes m. In general, there is a tradeoff between the repair bandwidth and the probability to using D2D download. How to set the parameter h depends on the frequency we can repair and code-rate for storage schemes in a DS system.

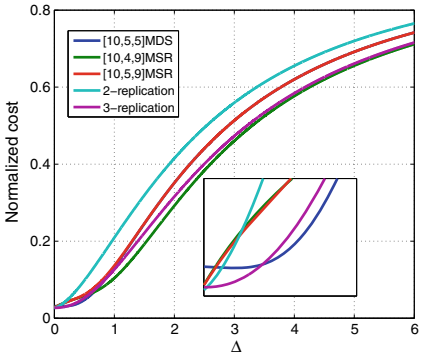


Fig. 5. Normalized cost various Δ for hybrid download and repair.

For a finite cost ratio ρ, the condition of reducing $\bar{C}_r$ is $\beta < \alpha$ and $\bar{C}_d$ is improved only if $\alpha < F$. So, we can know that both $\bar{C}_r$ and $\bar{C}_d$ are improved for MSR codes, and only $\bar{C}_d$ is reduced for MDS codes, however, there is no improvement for replication scheme, when using hybrid scheme. As shown in Fig. 5, it can be observed that MDS cod and MSR code both achieve a great improvement and they outperform replication scheme except instantaneous repair when using the hybrid scheme. That is because $\bar{C}_d$ is dominant to $\bar{C}$ and hybrid scheme reduces download probability from BS. However, as the download locality reduces, the coding rate $R = k/n$ is also decreased. Indeed, erasure codes are always performing the best for larger Δ, while replication performs worse in a DSS with D2D links.

6 Conclusions

In this paper, we investigated the impact of download locality and user number on communication cost in a wireless network with content cached in mobile devices. As the user increases, the communication cost reduces obviously. Moreover, reducing download locality can further reduce communication cost for lower repair frequency when using hybrid scheme. Particularly, MSR codes and MDS codes have equivalent cost for lower repair frequency and higher request frequency, however, MDS codes are less complex and simpler than MSR, and are more suitable for D2D system with distributed storage.

Acknowledgments. This work was supported by the National Natural Sciences Foundation of China under Grant 61701136, Grant 61501140, and Grant 61525103, and Project funded by China Postdoctoral Science Foundation 2018M630357, and Shenzhen Basic Research Program under Grant JCYJ20170811154233370, and ZDSYS201707280903305.

References

1. Wang, L., Wu, H., Han, Z.: Wireless distributed storage in socially enabled d2d communications. IEEE Access **4**, 1971–1984 (2017)
2. Li, Y., Wang, Z., Jin, D., Chen, S.: Optimal mobile content downloading in device-to-device communication underlaying cellular networks. IEEE Trans. Wirel. Commun. **13**(7), 3596–3608 (2014)
3. Golrezaei, N., Molisch, A.F., Dimakis, A.G.: Base-station assisted device-to-device communications for high-throughput wireless video networks. In: IEEE International Conference on Communications, pp. 7077–7081 (2013)
4. Golrezaei, N., Dimakis, A.G., Molisch, A.F.: Scaling behavior for device-to-device communications with distributed caching. IEEE Trans. Inf. Theory **60**(7), 4286–4298 (2012)
5. Paakkonen, J., Hollanti, C., Tirkkonen, O.: Device-to-device data storage for mobile cellular systems. In: GLOBECOM Workshops, pp. 671–676 (2013)
6. Pedersen, J., Amat, A.G.I., Andriyanova, I., Brännström, F.: Distributed storage in mobile wireless networks with device-to-device communication. IEEE Trans. Commun. **64**(11), 4862–4878 (2016)
7. Rashmi, K.V., Shah, N.B., Kumar, P.V.: Optimal exact-regenerating codes for distributed storage at the MSR and MBR points via a product-matrix construction. IEEE Trans. Inf. Theory **57**(8), 5227–5239 (2011)

Improved Precoding Scheme for MmWave Large-Scale MIMO Systems

Yiqi Lu, Chonghu Cheng(✉), Jie Yang, and Guan Gui

Nanjing University of Posts and Telecommunications, Nanjing 210003, China
chengch@njupt.edu.cn

Abstract. Millimeter wave (mmWave) system has been considered as a key technology in 5G, due to its large available bandwidth and high spectrum efficiency. For such system, the traditional full digital precoding is unrealistic due to the costs and complexity of radio-frequency(RF) chains. To overcome this difficulty, hybrid precoding is proposed. However, most existing hybrid precoding still involves high energy consumption. In this paper, we propose an improved hybrid precoding scheme based on energy-efficient switch and inverter (SI)-based precoding architecture. The improved hybrid precoding is aimed at adaptively updated the probability distribution of elements in hybrid precoder by minimizing cross entropy(CE), and reach a higher achievable sum rate by weighting. Simulation results prove that the proposed improved hybrid precoding scheme performs better than before.

Keywords: Hybrid precoding · Massive MIMO · mmWave · Adaptively update

1 Introduction

With the rapid development of mobile communication, because of the limited spectrum resources, the use of Millimeter wave (mmWave) with huge spectrum band from 30 GHz to 300 GHz has become a research hotspot. In mmWave communication system, the shorter wavelength of mmWave makes the antenna size smaller, so that large-scale antenna arrays can be deployed in the limited physical space in both base stations (BSs) and users, which is called massive multiple-input and multiple-output (MIMO), to enhance the spatial multiplexing gain or multiple access ability [1]. In addition, the application of massive MIMO provides sufficient array gain, to compensate for the severe path loss at this extremely high frequency band [2]. MmWave massive MIMO is regarded as a key technology in 5G [3]. In the MIMO system, a fully digital precoding architecture is adopted, in which each antenna needs a dedicated high energy consumption RF chain [4], and if this architecture is directly used in a mmWave massive MIMO system, it can lead to unacceptable hardware complexity and power consumption with the increase of the number of antennas. In order to reduce the number of RF chains, hybrid precoding is proposed [4], which is composed of a small-size digital baseband precoder and a large-size analog beamformer connected in parallel via a small number of RF chains [4]. However, most of the existing conventional hybrid precoding architectures require a large number of high-resolution phase shifters

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 385–393, 2019.
https://doi.org/10.1007/978-981-13-6264-4_47

in analog beamformer [5], which still involves high energy consumption. In [6], the authors propose an energy-efficiency switch and inverter (SI)-based precoding scheme, using a small number of energy-saving switches and inverters to replace high-energy phase shifters.

In this paper, the ACE algorithm in [6] is optimized based on a switch and inverter (SI)-based precoding architecture, using relative sum rate instead of the absolute sum rate to adaptively weight the hybrid precoders, to further improve the performance of sum rate and energy efficiency. Due to the adoption of the SI structure, the scheme we proposed also has the advantage of low energy consumption.

2 System Model

In this paper, we consider a multi-user downlink mmWave massive MIMO system, in which the base station (BS) is equipped with N transmit antennas and N_{RF} RF chains, serving K single-antennas users simultaneously. Generally, we have $K \leq N_{RF} \leq N$, to fully achieve the multiplexing gains, we assume $K = N_{RF}$ [7].

The transmit signal vector for all K users, denoted by $\mathbf{s} = [s_1, \ldots, s_K]^T$ with $\mathrm{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_K$, is processed in two steps with an $N_{RF} \times K$ digital baseband precoder F_{BB} followed by an $N \times N_{RF}$ analog beamformer $\boldsymbol{F}_{RF}$ realized by an analog circuit. The $K \times 1$ received signal vector $\mathbf{y}$ for all K users can be presented by

$$\mathbf{y} = \mathbf{H}\boldsymbol{F}_{RF}\boldsymbol{F}_{BB}\mathbf{s} + \mathbf{n} \tag{1}$$

where $\mathbf{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{h}_3, \ldots \boldsymbol{h}_K]^H$ is the channel matrix with $\boldsymbol{h}_k$ presenting the $N \times 1$ channel vector between the BS and kth user. In addition, $\boldsymbol{F}_{BB}$ and $\boldsymbol{F}_{RF}$ satisfy the total transmit power constraint as $\| \boldsymbol{F}_{RF}\boldsymbol{F}_{BB} \|_F^2 = \rho$. Finally, $\mathbf{n}$ is the additive white Gaussian noise vector of size $K \times 1$. When it comes to uniform planar array (UPA) with N_1 elements in horizon and N_2 elements in vertical constrained by $N = N_1 \times N_2$, the channel vector between the BS and kth user is given by [8]

$$\boldsymbol{h}_k = \sqrt{\frac{N_1 N_2}{L_k}} \sum\nolimits_{l=1}^{L_k} \alpha_k^l \mathbf{a}(\theta_k^l, \Phi_k^l) \tag{2}$$

where $\theta_k^l, \Phi_k^l \in (-\pi, \pi]$ denote horizontal azimuth and elevation angle, respectively, following the uniform distribution $\mathcal{U}(-\pi, \pi)$ [9], associated with the lth path. L_k represents the number of paths for the kth user. α_k^l is the complex gain of the lth path for user k. $\mathbf{a}(\theta, \Phi)$ is the array steering vector, which is a function of the antenna array structure. For $\sqrt{N} \times \sqrt{N}$ array dimension, $\mathbf{a}(\theta, \Phi)$ is given by

$$\begin{aligned} \mathbf{a}(\theta, \Phi) = [1, \cdots, e^{-j\frac{2\pi}{\lambda}d(m\sin(\theta)\sin(\Phi) + n\cos(\Phi))}, \\ \cdots, e^{-j\frac{2\pi}{\lambda}d\left((\sqrt{N}-1)\sin(\theta)\sin(\Phi) + (\sqrt{N}-1)\cos(\Phi)\right)}]^T \end{aligned} \tag{3}$$

where $0 \leq m \leq \sqrt{N}$, $0 \leq n \leq \sqrt{N}$. λ denotes carrier wavelength, and d is the antenna element spacing which usually be set to $\lambda/2$ [3].

3 Hybrid Precoding Scheme

In this section, we first introduce the traditional hybrid precoding architectures. Then, we introduce SI-based hybrid precoding architecture. Finally, we proposed an ACE with relative weight-based hybrid precoding scheme for SI-based architecture.

3.1 Traditional Hybrid Precoding Architecture

Most hybrid precoding architectures require a complicated high-resolution phase shifters network in the analogy beamformer. In order to solve the problem of high energy consumption that still exists, currently there are two solutions:

One is to replace the high-resolution phase shifters with finite-resolution phase shifters directly (PS-based architecture) [7, 10], shown in Fig. 1, and this architecture can reduce the power consumption of the phase shifter network without significant performance loss, achieving near-optimal performance. But it still requires $N \times N_{RF}$ finite-resolution phase shifters with considerable energy consumption expressed as

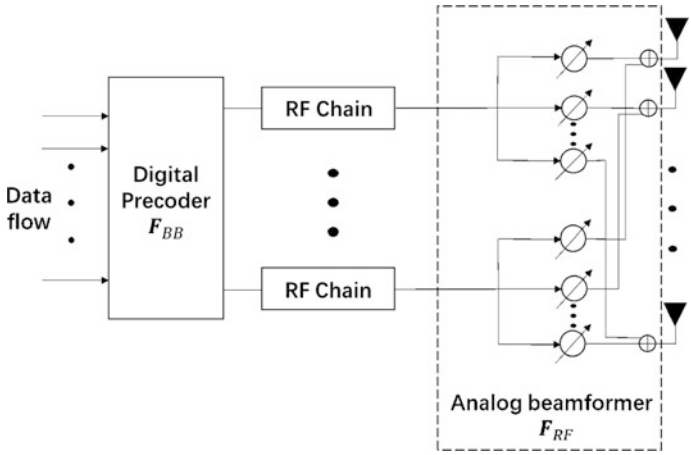


Fig. 1. Traditional PS-based architecture

$$P_{PS} = \rho + N_{RF}P_{RF} + NN_{RF}P_{PS} + P_{BB} \tag{4}$$

where $\| \boldsymbol{F}_{RF}\boldsymbol{F}_{BB} \|_F^2 = \rho$. P_{RF},P_{PS},P_{BB} represent the energy consumption of RF chain, finite-resolution phase shifter, and baseband, respectively. In this paper, we set $\rho =$ 30 mW [12], $P_{RF} = 300$ mW [12], $P_{BB} = 200$ mW [11], and $P_{PS} = 40$ mW (4-bit phase shifter) [11].

The other architecture shown in Fig. 2 is to replace phase shifter network with switch network(SW-based architecture) [10], which can significantly reduce the energy consumption. However, because only N_{RF} antennas work simultaneously, SW-based

architecture cannot fully realize the array gain, resulting in significant performance loss [6]. The energy consumption of SW-based architecture can be expressed as

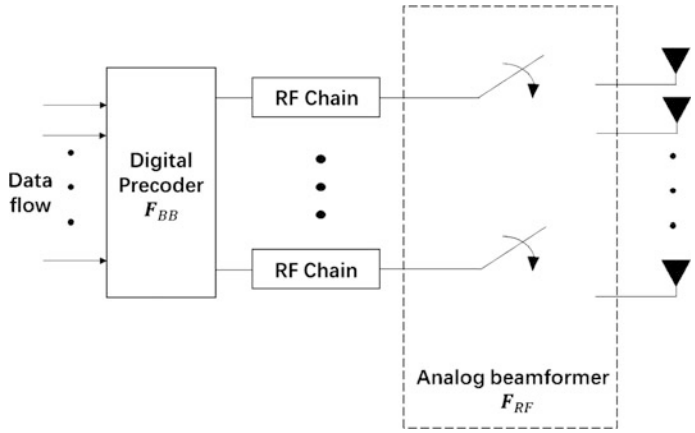


Fig. 2. Traditional SW-based architecture

$$P_{SW} = \rho + N_{RF}P_{RF} + N_{RF}P_{SW} + P_{BB} \tag{5}$$

And in this paper, we set $P_{SW} = 5$ Mw [11].

3.2 SI-Based Hybrid Precoding Architecture

The author in [6] proposed a compromise between two traditional precoding architectures: SI-based precoding structure, shown in the Fig. 3, in which each RF chain is connected to an antenna sub-array with only M antennas ($M = N/N_{RF}$ assumed to be an integer). Each antenna has a separate switch, so there is a higher array gain compared to SW-based architectures. In addition, it is proved in [6] that the loss of array gains keeps constant and limited when the number of BS antennas goes to infinity. The entire SI-based precoding architecture uses N switches and N_{RF} inverters instead of $N \times N_{RF}$ phase shifters. Compared to PS-based architectures, the energy consumption is greatly reduced, which can be expressed as

$$P_{SI} = \rho + N_{RF}P_{RF} + N_{RF}P_{IN} + NP_{SW} + P_{BB} \tag{6}$$

where P_{IN} is the energy consumption of inverter, which is similar to the energy consumption of switches [10] considered as $P_{IN} = P_{SW} = 5$ mW [11]. In addition, the energy consumption for fully digital ZF precoding should be

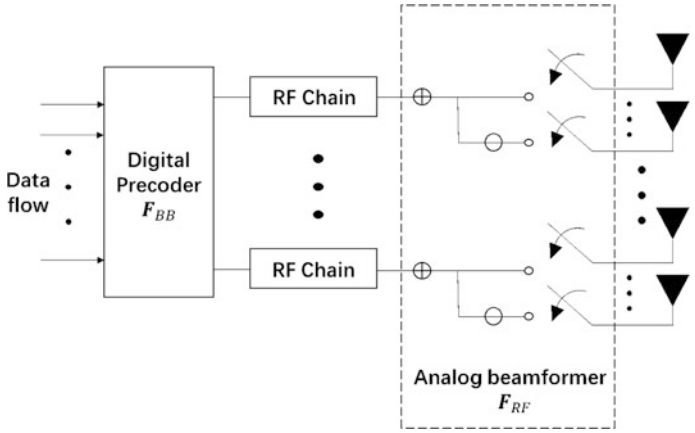


Fig. 3. SI-based architecture

$$P_{ZF} = \rho + NP_{RF} + P_{BB} \tag{7}$$

The SI-based architecture has two hardware constraints. First, the analog beamformer F_{RF} should be a block diagonal matrix with the size of $N \times N_{RF}$ shown by

$$\boldsymbol{F}_{RF} = \begin{bmatrix} \boldsymbol{F}_1^{RF} & 0 & \dots & 0 \\ 0 & \boldsymbol{F}_2^{RF} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \boldsymbol{F}_{N_{RF}}^{RF} \end{bmatrix} \tag{8}$$

where F_n^{RF} is the $M \times 1$ analog precoding matrix of the nth sub antenna array. Second, all the N nonzero elements of $\boldsymbol{F}_{RF}$ should belong to the set

$$\{-1, +1\} \tag{9}$$

due to only inverters and switches are used.

3.3 ACE with Relative Weight-Based Precoding Base on SI-Based Architecture

Our purpose is to design $\boldsymbol{F}_{RF}$ and $\boldsymbol{F}_{BB}$, to maximize the achievable sum rate R. This can be described as

$$\left(\boldsymbol{F}_{RF}^{opt}, \boldsymbol{F}_{BB}^{opt}\right) = \arg \max_{\boldsymbol{F}_{RF}, \boldsymbol{F}_{BB}} \mathrm{R} \qquad \text{s.t.} \boldsymbol{F}_{RF} \in \mathbb{F}, \| \boldsymbol{F}_{RF}\boldsymbol{F}_{BB} \|_F^2 = \rho \tag{10}$$

where $\mathbb{F}$ is the set of all the possible $\boldsymbol{F}_{RF}$ which satisfy the two constraints (8) and (9), and the formula of sum rate R is given below

$$\mathrm{R} = \sum\nolimits_{k=1}^{K} \log_2 \left(1 + \frac{\left| \boldsymbol{h}_k^H \boldsymbol{F}_{RF} f_k^{BB} \right|^2}{\sum_{k' \neq k}^{K} \left| \boldsymbol{h}_k^H \boldsymbol{F}_{RF} f_{k'}^{BB} \right|^2 + \sigma^2} \right) \tag{11}$$

where f_k^{BB} denotes the kth columns of $\boldsymbol{F}_{BB}$. The specific steps of the proposed algorithm are as follows:

Step 1, we use f_j as the value of jth nonzero element, $u_j (1 \leq j \leq N)$ as the probability of $f_j = +1$, and $1 - u_j$ as the probability of $f_j = -1$, so the probability of N nonzero elements in (8) is expressed as $\mathbf{u} = [u_1, u_2, u_3 \ldots u_N]^T$, where $[u_{(n-1)\times M+1}, u_{(n-1)\times M+2}, \ldots, u_{n\times M}]$ stands for the probability of M nonzero elements in $\boldsymbol{F}_n^{RF}$ $(1 \leq n \leq N_{RF})$. Since there is no prior information, the probability is initialized to $\mathbf{u}^{(0)} = (1/2) \times 1_{N\times 1}$ ($1_{N\times 1}$ is a full 1 vector).

Step 2, according to the probability distribution $u^{(m)}$, S candidate $\boldsymbol{F}_{RF}^i (1 \leq \mathrm{i} \leq \mathrm{S})$ are randomly generated, and m represents the m th iteration.

Step 3, calculate corresponding $\boldsymbol{F}_{BB}^i$ $(1 \leq \mathrm{i} \leq \mathrm{S})$ by

$$\boldsymbol{F}_{BB}^i = \frac{\sqrt{\rho}}{||\boldsymbol{F}_{RF}^i((\mathbf{H}_{eq}^i)^H(\mathbf{H}_{eq}^i(\mathbf{H}_{eq}^i)^H)^{-1})||_F} (\mathbf{H}_{eq}^i)^H (\mathbf{H}_{eq}^i(\mathbf{H}_{eq}^i)^H)^{-1} \tag{12}$$

where $\mathbf{H}_{eq}^i$ denotes effective channel and we have $\mathbf{H}_{eq}^i = \mathbf{H}\boldsymbol{F}_{RF}^i$, then calculate sum rate $\left\{\mathrm{R}(\boldsymbol{F}_{RF}^i)\right\}_{i=1}^{\mathrm{S}}$ by (11).

Step 4, sort $\left\{\mathrm{R}(\boldsymbol{F}_{RF}^i)\right\}_{i=1}^{\mathrm{S}}$ in descending order as $\mathrm{R}(\boldsymbol{F}_{RF}^1) \geq \mathrm{R}(\boldsymbol{F}_{RF}^2) \geq \ldots \geq \mathrm{R}(\boldsymbol{F}_{RF}^S)$.

Step 5, select the former S_{elite} $\boldsymbol{F}_{RF}^i$ as elites $\boldsymbol{F}_{RF}^1, \boldsymbol{F}_{RF}^2 \ldots \boldsymbol{F}_{RF}^{\mathrm{S}_{elite}}$.

Step 6, calculate the weights of elite $\boldsymbol{F}_{RF}^i$

$$\omega_i = \frac{\mathrm{S}_{elite} |\mathrm{R}(\boldsymbol{F}_{RF}^i) - \mathrm{R}(\boldsymbol{F}_{RF}^{\mathrm{S}_{elite}})|}{\sum_{i=1}^{\mathrm{S}_{elite}} |\mathrm{R}(\boldsymbol{F}_{RF}^{\mathrm{i}}) - \mathrm{R}(\boldsymbol{F}_{RF}^{\mathrm{S}_{elite}})|} \qquad (1 \leq \mathrm{i} \leq \mathrm{S}_{elite}) \tag{13}$$

Step 7, update $\boldsymbol{u}^{(m+1)}$ with elites by minimizing CE

$$\boldsymbol{u}^{(m+1)} = \arg\max_{u^{(m)}} \frac{1}{\mathrm{S}} \sum_{i=1}^{\mathrm{S}_{elite}} \omega_i \ln \Xi (\boldsymbol{F}_{RF}^i; \boldsymbol{u}^{(m)}) \tag{14}$$

where $\Xi(\boldsymbol{F}_{RF}^i; \boldsymbol{u}^{(m)})$ denotes the probability distributions of generated $\boldsymbol{F}_{RF}^i$ with probability $\boldsymbol{u}^{(m)}$

$$\Xi\left(\boldsymbol{F}_{RF}^i; \boldsymbol{u}^{(m)}\right) = \prod_{j=1}^{N} (\boldsymbol{u}_j^{(m)})^{\frac{1}{2}\left(1+f_j\right)} (1 - \boldsymbol{u}_j^{(m)})^{\frac{1}{2}\left(1-f_j\right)} \tag{15}$$

Step 8, find the first derivative of $\frac{1}{S}\sum_{i=1}^{S_{elite}} \omega_i \ln \Xi(\boldsymbol{F}_{RF}^{i}; \boldsymbol{u}^{(m)})$ and make it zero to solve $\boldsymbol{u}^{(m)}$ as updated $u^{(m+1)}$

$$\boldsymbol{u}_j^{(m+1)} = \frac{\sum_{i=1}^{S_{elite}} \omega_i (f_j + 1)}{2 \sum_{i=1}^{S_{elite}} \omega_i} \tag{16}$$

Return to Step 2 and make $m = m + 1$ to repeat this process. This iteration repeats until m reaches the maximum number of iterations. Finally, $\boldsymbol{F}_{RF}^{1}$ and $\boldsymbol{F}_{BB}^{1}$ are taken as optimal analog beamformer $\boldsymbol{F}_{RF}^{opt}$ and optimal digital precoder $\boldsymbol{F}_{BB}^{opt}$.

4 Simulation Results and Discussion

In this section, we provide some simulation results for the comparison to prove the superiority of our proposal in terms of achievable sum-rate performance. The mmWave channel between the BS and kth user is generated according to the channel model in (2) with $L_k = 3$ scatterers and $\alpha_k^l \sim \mathcal{CN}(0, 1)$ for $1 \leq l \leq L_k$. We assume that the BS employs a UPA with antenna spacing $= \lambda/2$. Finally, the signal-to-noise ratio (SNR) is defined as ρ/σ^2.

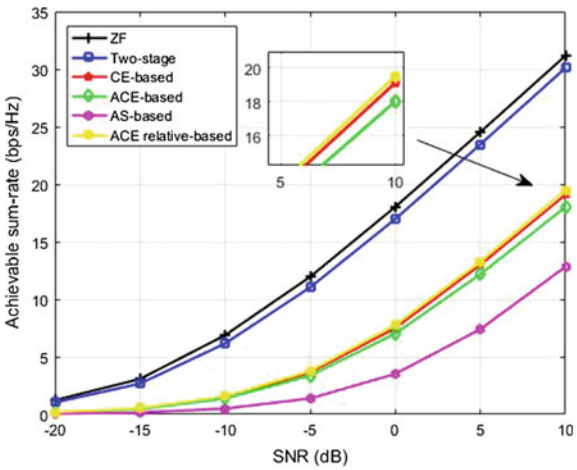


Fig. 4. Achievable sum-rate comparison

Figure 4 shows the achievable sum rate comparison of different schemes in a mmWave massive MIMO system with $N = 128$, $K = N_{RF} = 4$. CE-based using conventional CE algorithm, ACE-based using the ACE algorithm proposed in [6], and ACE with relative weight based hybrid precoding we proposed are designed for SI-based architecture and the parameters are set as follows: $S = 200$, $S_{elite} = 40$, and the number of iterations $I = 16$, while the traditional two-stage hybrid precoding is

designed for PS architecture with 4-bit phase shifters [7] and the antenna selection (AS)-based hybrid precoding is used for switches(SW)-based architecture [10]. As we observe from Fig. 5, the proposed algorithm is superior to ACE algorithm with stable promotion against SNR. Also, we notice that the proposed hybrid precoding can achieve much higher achievable sum rate than AS-based hybrid precoding, which means obtaining higher array gains in mmWave massive MIMO system.

Figure 5 depicts the performance comparison of algorithm before and after improvement against the number of candidates and the number of iterations, when SNR = 10 dB, $N = 128$, $\mathrm{K} = N_{RF} = 4$, and $\mathrm{S}_{elite}/\mathrm{S} = 0.2$. First, simulation result shows that the sum rate performance of the improved algorithm is significantly better than that of the ACE algorithm under the same conditions. Second, when the number of iterations reaches 16, the available sum rate of proposed algorithm and the performance gap both tend to be stable, so the number of iterations does not have to be too large, and 16 is sufficient.

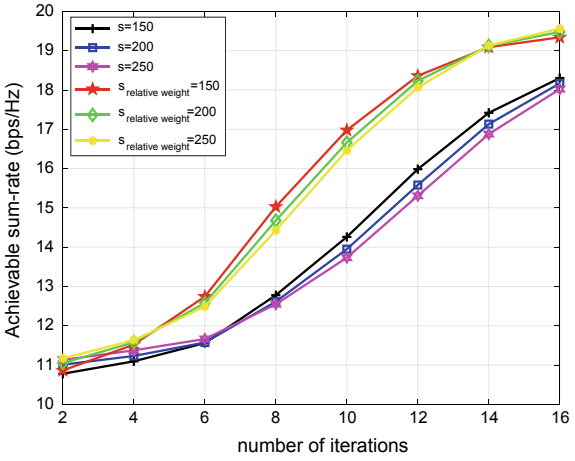


Fig. 5. Achievable sum-rate comparison of algorithm before and after improvement

5 Conclusion

In this paper, we proposed an improved ACE-based precoding with relative weight using SI-based hybrid precoding architecture. The simulation result proves that the precoding scheme we proposed can achieve higher sum rate while guaranteeing energy efficiency compared to the conventional ACE-based precoding scheme.

References

1. Larsson, E., Edfors, O., Tufvesson, F., Marzetta, T.: Massive MIMO for next generation wireless systems. IEEE Commun. Mag. **52**(2), 186–195 (2014)
2. Rangan, S., Rappaport, T.S., Erkip, E.: Millimeter-wave cellular wireless networks: potentials and challenges. Proc. IEEE **102**(3), 366–385 (2014)

3. For, O.: Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G, no. January, pp. 186–194, 2015
4. El Ayach, O., Rajagopal, S., Abu-Surra, S., Pi, Z., Heath, R.W.: Spatially sparse precoding in millimeter wave MIMO systems. IEEE Trans. Wirel. Commun. **13**(3), 1499–1513 (2014)
5. Chen, J.-C.: Energy-efficient hybrid precoding design for millimeter-wave massive MIMO systems via coordinate update algorithms. IEEE Access, vol. 3536, no. c, pp. 1–1, 2018
6. Gao, X., Dai, L., Sun, Y., Han, S., Chih-Lin, I.: Machine learning inspired energy-efficient hybrid precoding for mmWave massive MIMO systems. IEEE Int. Conf. Commun. (2017)
7. Alkhateeb, A., Member, S., Leus, G., Heath, R.W.: Limited feedback hybrid precoding for multi-user millimeter wave systems, vol. 14, no. 11, pp. 6481–6494 (2015)
8. Heath Jr, R.W., González-prelcic, N., Rangan, S., Roh, W., Sayeed, A.M.: An overview of signal processing techniques for millimeter wave MIMO systems, vol. 10, no. 3, pp. 436–453 (2016)
9. Sayeed, A., Brady, J.: Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies. In: GLOBECOM-IEEE Global Telecommunications Conference, pp. 3679–3684 (2013)
10. Mendez-rial, R., Rusut, C., Alkhateeb, A., Gonzalez-prelcict, N., Heath Jr, R.W.: Channel estimation and hybrid combining for mmWave: phase shifters or switches?
11. Méndez-rial, R.O.I., Rusu, C., González-prelcic, N., Alkhateeb, A., Member, S., Heath, R. W.: Hybrid MIMO architectures for millimeter wave communications: phase shifters or switches? vol. 4, pp. 247–267 (2016)
12. Gao, X., Dai, L., Han, S., Chih-Lin, I., Heath, R.W.: Energy-efficient hybrid analog and digital precoding for MmWave MIMO systems with large antenna arrays. IEEE J. Sel. Areas Commun. **34**(4), 998–1009 (2016)

Soft-Decision Multiple Access TDCS and PAPR Reduction

Su Hu[1,2(✉)], Fan Li[1], Shiyong Ma[1], and Yuan Gao[3,4(✉)]

[1] National Key Laboratory on Communications,
University of Electronic Science and Technology of China, Chengdu 611731, China
[2] The Center for Intelligent Networking and Communication,
University of Electronic Science and Technology of China, Chengdu 611731, China
husu@uestc.edu.cn
[3] Academy of Military Science of PLA, Beijing 100091, China
[4] Department of Electronic Engineering, Tsinghua University,
Beijing 100084, China
284480150@qq.com

Abstract. As a candidate of overlay cognitive radio (CR), transform domain communication system (TDCS) can achieve dynamic spectrum access by avoiding existing users and interference spectrum bins. However, in poor channel conditions, traditional TDCS cannot maintain low probability of interception (LPI) and covert communication capability at the same time. A TDCS structure of designing the fundamental modulation waveform (FMW) with a soft-decision (SD) method is presented in this paper. To reduce peak-to-average power ratio (PAPR) of SD-TDCS, an algorithm which designs the low-sidelobe FMWs under spectrum constrains is proposed. Simulation results demonstrate that the proposed SD-TDCS architecture can improve BER performance and anti-interference ability. And adequate PAPR reduction is achieved with the proposed waveform design algorithm.

Keywords: Transform-domain communication system · Soft decision · PAPR reduction

1 Introduction

Owing to the increase of wireless communication applications and fixed spectrum assignment policy, spectrum becomes more and more congested [1]. The concept of cognitive radio (CR) is proposed as a solution to the spectrum scarcity problem [2]. As an overlay CR technique, transform domain communication system (TDCS) has attracted wide research interests in recent years [3,4].

TDCS transmits a specific cyclic shift version of fundamental modulation waveform (FMW) according to data input [3]. The design of FMW is the core of TDCS since it is closely related to the anti-jamming ability and communication performance of the system. Traditional TDCS uses 40% of the max spectrum amplitude as the threshold based on numerous simulation experiments [5]. Fixed

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 394–402, 2019.
https://doi.org/10.1007/978-981-13-6264-4_48

threshold is sensitive to noise power and cannot adapt to different environments flexibly. To deal with this problem, an improved structure of soft-decision (SD) TDCS is proposed and studied in this paper.

Furthermore, TDCS is a multicarrier system, whose time-domain sample is a weighted sum of independent random variables with unit magnitude and a pseudorandom phase [4]. This leads to a high peak-to-average power ratio (PAPR) of transmitting waveforms, which makes TDCS particularly sensitive to nonlinear distortion caused by high power amplifier (HPA). Numerous algorithms have been proposed to deal with PAPR problems in multi-carrier systems, including tone reservation (TR) [6], clipping [7], probability [8] and coding [9]. However, the clipping and coding methods are not appropriate for TDCS. Probability and TR methods require high computational complexity and side information (SI). In this paper, a PAPR reduction algorithm based on waveform design is proposed. It is observed that the proposed algorithm can efficiently reduce PAPR of multiuser SD-TDCS under spectrum constrains.

2 Review of Traditional Multiuser TDCS

The block diagram of multiuser TDCS transmitter is illustrated in Fig. 1. In TDCS, the whole spectrum is divided into N spectrum bins. After this, a threshold is applied to avoid interference frequency bins. Assume all users are in the same environment, a spectrum marking vector, $\mathbf{S}$ is generated to indicate the status of spectrum bins. The value of S_k is set to 1 (or 0) if the spectrum estimation results of the k-th subcarrier is smaller (or larger) than the threshold.

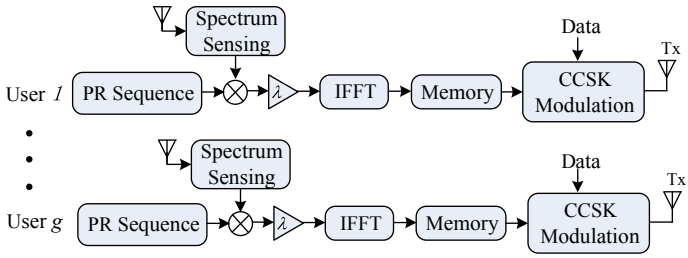


Fig. 1. Block diagram of traditional multiuser TDCS transmitter.

For simplification, we take the g-th user for example. After spectrum estimation and marking, a multi-valued complex pseudorandom (PR) phase vector $\mathbf{P}_g$ is generated and multiplied element-by-element with $\mathbf{S}$ to produce a frequency-domain FMW $\mathbf{B}_g$:

$$\mathbf{B}_g = \mathbf{S} \cdot \mathbf{P}_g = \{S_0 e_g^{jm_0}, S_1 e_g^{jm_1}, \cdots, S_{N-1} e_g^{jm_{N-1}}\}, \tag{1}$$

In multiple access environments, by allocating users different PR sequences the pseudo-orthogonality between different FMWs can be realized so as to achieve multiuser access for TDCS. The frequency-domain FMW is amplitude-scaled to ensure all symbols are transmitted with equal energy and then inverse discrete Fourier transform (IDFT) is applied to produce a time-domain FMW:

$$\mathbf{b}_g = \mathrm{IDFT}\left\{\mathbf{B}_g\right\} \Leftrightarrow b_g^n = \frac{1}{N}\sum_{k=0}^{N-1} B_g^k e^{j\frac{2\pi kn}{N}} = \frac{\lambda}{N}\sum_{k=0}^{N-1} S_k e_g^{jm_k} e^{\frac{j2\pi kn}{N}}. \quad (2)$$

where λ is the scaling factor.

Following IDFT modulation, $\mathbf{b}_g$ is modulated using CCSK. The modulated TDCS waveform $\mathbf{x}_g$ is expressed by

$$\mathbf{x}_g = \left\{x_g^0, x_g^1, \cdots, x_g^{N-1}\right\} = \left\langle \mathbf{b}_g \right\rangle_{d_i}. \quad (3)$$

where d_i is the ith transmitting data, M_{ary} is the CCSK modulation order, and $\langle\cdot\rangle_d$ denotes shifting $\mathbf{b}_g$ cyclically to the left by d places.

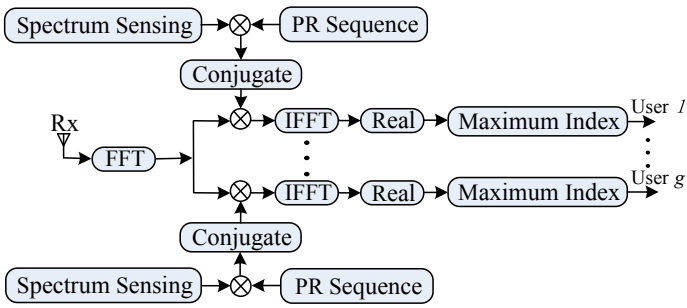


Fig. 2. Block diagram of traditional multiuser TDCS receiver.

The block diagram of multiuser TDCS receiver is illustrated in Fig. 2. Assume the transmitter and receiver get identical spectrum sensing results. The received signal $\mathbf{r} = \{r_0, r_1, \cdots, r_{N-1}\}$ is correlated with local FMW signal $\mathbf{b}_g$. Thus, the input data $\mathbf{d}$ can be recovered by the location of the maximum correlation value. To eliminate noise effects, only the real part is utilized:

$$\tilde{\mathbf{d}} = \arg\max\{\Re\{\mathcal{F}^{-1}\{\mathcal{F}(\mathbf{r}) \cdot (\mathcal{F}(\mathbf{b}_g))*\}\}\}. \quad (4)$$

where $\Re\{\cdot\}$ denotes obtaining the real component of a complex number, $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ denote operations of DFT and its inverse (IDFT), respectively.

3 SD-TDCS Structure

3.1 Design of SD-TDCS

In poor channel conditions, traditional hard-decision TDCS has some problems:

(1) If the threshold is set too low, most interference spectrum bins are unavailable and the correlation of FMW are damaged, which results in poor Bit Error Rate (BER) performance. Besides, when the transmission power is unchanged, the power of each spectrum bin is very high, which damages covert communication ability.

(2) If the threshold is set too high, the frequencies with interference may not be avoided, which results in a higher BER and poor anti-interference ability.

To maintain anti-interference and covert communication ability at the same time, this article proposes a soft-decision method to mark the spectrum. Spectrum estimation value is denoted as I_k. The spectrum whose amplitude is higher than the high threshold T_{high} is marked as 0, indicating that it is unavailable. The spectrum below the low threshold T_{low} is marked as 1 to indicate that it is available. Spectrum that lies between the two thresholds is marked as

$$S_k = \frac{I_k - T_{\text{low}}}{T_{\text{high}} - T_{\text{low}}}. \tag{5}$$

The availability of the spectrum bins is represented by variable **S** with the range changed from binary numbers (hard decision) to real values between 0 and 1. Furthermore, **S** also represents the allowable transmission power densities, i.e., the spectrum bins suffering from higher interference are allocated less energy to transmit data and vice versa. The FMW obtained in this way can reflect the actual electromagnetic environment, as shown in Fig. 3.

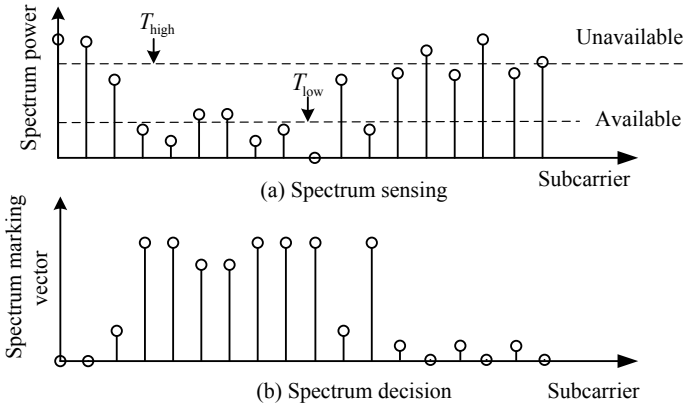


Fig. 3. Spectrum sensing results of SD-TDCS.

3.2 PAPR Reduction of TDCS

The instantaneous power of the time-domain signal $x(t)$ is directly related to the sidelobes of autocorrelation function of frequency-domain signal X_n [10]. Since the transmitting signal of TDCS is a circshift version of time-domain FMW, we can design a frequency-domain FMW with low integrated sidelobe level of periodic autocorrelation to reduce PAPR.

For SD-TDCS, the FMW is a product of PR sequence and spectrum marking vector. Denote spectrum marking vector as $\mathbf{S}$ and new PR sequence as $\mathbf{Z}$. FMW in frequency domain is $\mathbf{X}$=$\mathbf{S} \cdot \mathbf{Z}$. The autocorrelation of waveform $\mathbf{X}$ can be written as

$$r_k = \sum_{n=0}^{N-1} X_n {X^*}_{(n+k)} = {r^*}_{-k}, k = 0, 1, \cdots, N-1. \tag{6}$$

Let $\mathbf{M}$ denote the following $N \times N$ right-circulant matrix:

$$\mathbf{M} = \begin{bmatrix} X_0 & X_1 \cdots X_{N-1} \\ X_{N-1} & X_0 \cdots X_{N-2} \\ \vdots & \vdots \ddots \vdots \\ X_1 & X_2 \cdots X_0 \end{bmatrix}, \tag{7}$$

Then, $N \times N$ correlation matrix of FMW can be written as

$$\begin{bmatrix} r_0 & {r_1}^* & {r_{N-1}}^* \\ r_1 & r_0 & {r_{N-2}}^* \\ \vdots & \vdots \ddots \vdots \\ r_{N-1} & r_{N-2} \cdots r_0 \end{bmatrix} = \mathbf{M}\mathbf{M}^H. \tag{8}$$

Therefore, we can design the desired waveform with impulse-like periodic correlations by minimizing the following criterion [11]:

$$C_1 = \left\| \mathbf{M}\mathbf{M}^H - r_0 \mathbf{I}_N \right\|^2. \tag{9}$$

Then the criterion can be rewritten as

$$C_1 = \sum_{\substack{k=-(N-1) \\ k \neq 0}}^{N-1} (N - |k|) |r_k|^2 = 2 \sum_{k=1}^{N-1} (N-k) \left| r_k \right|^2. \tag{10}$$

It is observed from (10) that the larger correlation lags have larger weights. Following this criterion, a more natural criterion where all correlation lags have equal weights is expressed as

$$C_2 = \sum_{k=1}^{N-1} N \left| r_k \right|^2. \tag{11}$$

In the periodic case, the two criteria are essentially identical [11]. Let $\mathbf{f}$ denote $N \times N$ IDFT matrix and the (k, p)-th element is

$$f_{(k,p)} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}kp} \; k, p = 0, 1, \cdots, N-1, \tag{12}$$

The IDFT of sequence $\{X_n\}$ is denoted as

$$x_p = \sum_{n=0}^{N-1} X_n e^{j\frac{2\pi}{N}np} \; p = 0, 1, \cdots, N-1, \tag{13}$$

Then it is easy to know:

$$\mathbf{x} = \mathbf{f}^H \mathbf{df}, \tag{14}$$

where

$$\mathbf{d} = \begin{bmatrix} x_1 & & 0 \\ & \ddots & \\ 0 & & x_N \end{bmatrix}. \tag{15}$$

The criterion C_1 is a quartic function of waveform elements. A simpler quartic function related to C_1 is written as [11]:

$$C_3 = \|\mathbf{x} - \sqrt{r_0}\mathbf{U}\|^2, \tag{16}$$

where $\mathbf{U}$ is a unitary matrix with $N \times N$ elements.

Similarly, a simpler criterion related to (16) in the frequency domain is expressed as [11]

$$C_4 = \sum_{p=0}^{N-1} \left| x_p - \sqrt{r_0} e^{j\psi_p} \right|^2. \tag{17}$$

where ψ_p are auxiliary variables.

In the spectrum-constrained scenario, minimizing C_1 or C_2 is the best way for designing an optimal sequence. For computation simplicity, we choose to design a quasi-optimal waveform by minimizing the simpler criteria C_3 and C_4. Here, we choose C_4 since its computational complexity is smaller. Since the minimizing algorithm is an improvement on the basis of periodic cyclic algorithm-new (Pecan) algorithm under spectrum constrains, we call it spectrum-constrained Pecan (SC-Pecan) algorithm. The steps of SC-Pecan algorithm are illustrated in Table. 1:

Table 1. SC-Pecan iterative algorithm

Step 1 Use N independent and uniformly distributed phases in the interval $[0, 2\pi]$ to generate an initial sequence $\mathbf{Z}$. Multiply spectrum marking vector $\mathbf{S}$ and $\mathbf{Z}$ to get a spectrum-constrained initial sequence $\{X_n\}$.
Step 2 Compute the IDFT of $\mathbf{X}$, i.e, $\{x_p\}_{p=0}^{N-1}$. The minimization problem turns into minimizing the auxiliary variable: $\psi_p = \arg(x_p), p = 0, 1, \cdots, N-1$.
Step 3 For given ψ_p, denote the DFT of $\{e^{j\psi_p}\}$ as $\{y_n\}$. Then the minimizing $\mathbf{X}$ is given by: $X_n = e^{j \arg(y_n)} \cdot S_n, n = 0, 1, \cdots N-1$.
Step 4 Repeat Steps 2 and 3 until the preset stop criterion is satisfied.

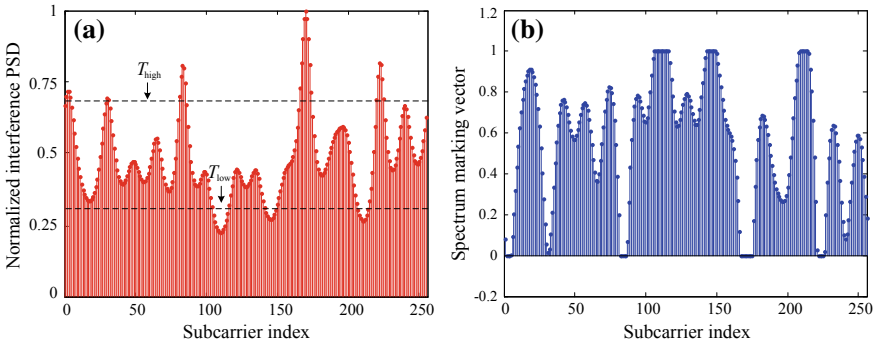


Fig. 4. The interference PSD and spectrum marking results.

4 Simulation Reuslts

In this section, some simulations are performed to verify the proposed methods and analytical results above.

The spectrum estimation is implemented with Burg's method, which is proposed in [12]. The estimation of a chirp interference is shown in Fig. 4a and spectrum marking results are shown in Fig. 4b. Here we choose 70% and 30% of the max spectrum amplitude as the high and low thresholds, respectively, which are considered to be optimal in numerical simulations. It is observed that the spectrum is heavily jammed, but most spectrum bins are available for transmission by limiting its transmission power.

We consider an SD-TDCS system employing N = 256 subcarriers, $Mary$ = 256-order CCSK modulation and 10dB signal-to-interference ratio. Assume the spectrum marking vectors of the transmitter and receiver are perfectly matched. Fig. 5 shows BER performance of SD-TDCS and traditional TDCS with different users in AWGN channel. The solid lines represent SD-TDCS and dotted

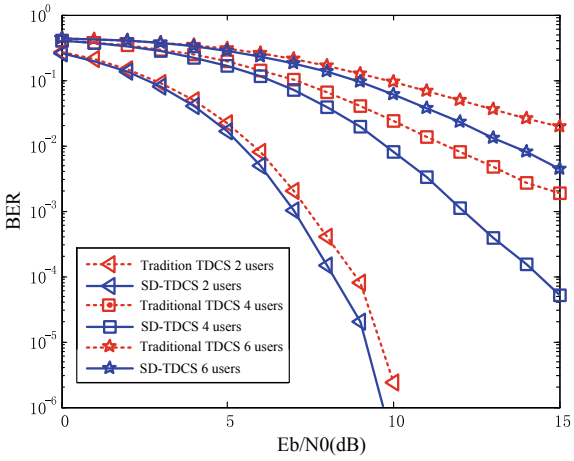


Fig. 5. BER performance of different users for SD-TDCS and traditional TDCS.

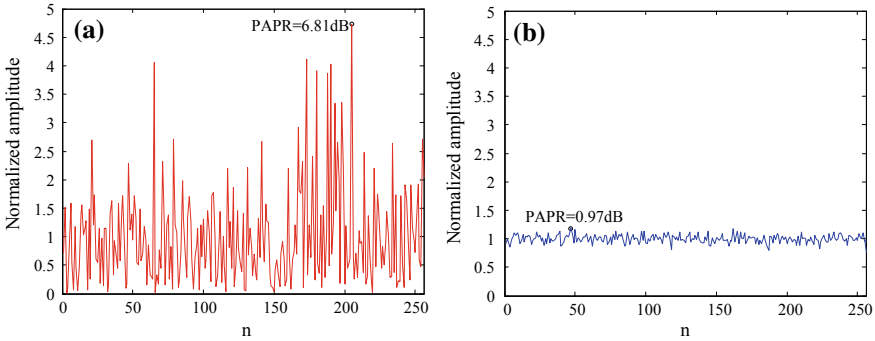


Fig. 6. Normalized amplitudes of transmitting signals.

lines represent traditional TDCS. It is seen that SD-TDCS has better BER performance than traditional TDCS. This is because when the interference covers a wide range of the spectrum, most spectra are unavailable for traditional TDCS, which results in poor autocorrelation of time-domain FMW and poor cross-correlations of FMWs for different users. While SD-TDCS maintains good correlation property since most spectra are allowable for transmission.

In Fig. 6, comparisons on normalized amplitudes of transmitting signals generated from PR sequence and SC-Pecan sequences are given. It can be seen that the signal generated from SC-Pecan sequence has small fluctuations, and the PAPR is 0.97 dB. While the normalized amplitude of the signal generated from PR sequence has extremely high peaks, and the PAPR is 6.81 dB.

5 Conclusion

To improve anti-interference performance and maintain convert communication capability of TDCS in poor channel conditions, we developed an SD-TDCS structure for multiple access TDCS. The proposed structure makes up the deficiency of traditional TDCS but still has a high PAPR. To reduce PAPR, we proposed an SC-Pecan waveform design method to optimize the correlation of FMW in the frequency domain. Simulation results show that the proposed system outperforms traditional TDCS in BER performance and the SC-Pecan algorithm can efficiently reduce PAPR of SD-TDCS under spectrum constrains.

References

1. Mitola, J.I., Maguire, G.Q.J.: Cognitive radio: making software radios more personal. IEEE Pers. Commun. **6**(4), 13–18 (1999)
2. Haykin, S.: Cognitive radio: brain-empowered wireless communications. IEEE J. Sel. Areas Commun. **23**(2), 201–220 (2005)
3. Chakravarthy, V., Nunez, A.S., Stephens, J.P., Shaw, A.K., Temple, M.A.: TDCS, OFDM, and MC-CDMA: a brief tutorial. IEEE Commun. Mag. **43**(9), S11–S16 (2005)
4. Hu, S., Bi, G.A., Guan, Y.L., Li, S.: TDCS-based cognitive radio networks with multiuser interference avoidance. IEEE Trans. Commun. **61**(12), 4828–4835 (2013)
5. Hu, S., Bi, G.A., Guan, Y.L., Li, S.: Spectrally efficient transform domain communication system with quadrature cyclic code shift keying. IET Commun. **7**(4), 382–390 (2013)
6. Tellado-Mourelo, J.: Peak to average power reduction for multicarrier modulation (1999)
7. Gong, L., Yang, S.H., Chen, Y.: Research on the reduction of PAPR for OFDM signals by companding and clipping method. WiCOM, pp. 1–4 (2010)
8. Xiao, Y., Wen, Q., Lei, X., Li, S.: Improved PTS for PAPR reduction in OFDM systems. IEEE ICT, p. 37 (2007)
9. Jones, A.E., Wilkinson, T.A., Barton, S.K.: Block coding scheme for reduction of peak to mean envelope power ratio of multicarrier transmission schemes. Electron. Lett. **30**(25), 2098–2099 (1994)
10. Xiao, Y., Lei, X., Wen, Q., Li, S.: A class of low complexity PTS techniques for PAPR reduction in OFDM systems. IEEE Signal Process. Lett. **14**(10), 680–683 (2007)
11. Stoica, P., He, H., Li, J.: New algorithms for designing unimodular sequences with good correlation properties. IEEE Trans. Signal Process. **57**(4), 1415–1425 (2009)
12. Emresoy, M.K., El-Jaroudi, A.: Evolutionary Burg spectral estimation. IEEE Signal Process. Lett. **4**(6), 173–175 (1997)

Development and Performance Evaluation of TDCS Multiple Access System

Su Hu[1,2(✉)], Qu Luo[1], Pei Wang[1], Zilong Liu[3], Dan Huang[1], and Yuan Gao[4]

[1] National Key Laboratory on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China[2] The Center for Intelligent Networking and Communications (CINC), University of Electronic Science and Technology of China (UESTC), Chengdu 611731, China
husu@uestc.edu.cn

[3] Institute of Communication Systems, 5G Innovation Centre, University of Surrey, GU2 7XH, United Kingdom[4] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

Abstract. In this paper, the transform-domain communication system based multi-user access system has been implemented and demonstrated via software-defined radio. A practical testbed with two National Instruments PXIe devices and four universal software defined radio reconfigurable input/output (USRP RIO) devices is presented for system implementation. A successive interference cancellation (SIC) receiver is designed to improve its multiuser performance. The testbed is capable of detecting interference and transmitting in real-time. Additionally, the centralized layout and distributed layout of system devices are both evaluated. We have demonstrated and demonstrated real-time seamless data transmission among six users, while avoiding environment interference. Details of system design including testbed synchronization, SIC receiver, and bit error rate are also presented.

Keywords: Transform-domain communication system
Universal software defined radio · Testbed setup
Successive interference cancelation

1 Introduction

Transform-domain communication system (TDCS) is considered to be a promising cognitive radio (CR) technology to provide the capability to use or share the spectrum [1–3]. TDCS can achieve reliable communication and low probability of interception (LPI) characteristic through frequency-domain spreading techniques. Furthermore, unlike orthogonal frequency division multiplexing (OFDM), TDCS was mainly developed for noncontinuous spectrum access and anti-interference transmission technology by spectrum nulling [4].

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 403–411, 2019.
https://doi.org/10.1007/978-981-13-6264-4_49

Software-defined radio (SDR) is one of the most evolving prototyping tools in industrial, academic areas, and commercial sectors. It is easy to set up a prototype to evaluate the system performance by occupying universal software radio peripheral (USRP) [5], which is a flexible and low-cost SDR platform [2]. Contributions about SDR implementation for multiple access system are very limited. In [6], authors have developed a practical non-orthogonal multiple access (NOMA) system based on USRP platform. In [7], a TDCS-based USRP platform is presented for control channels of cognitive radio networks (CRNs).

Extensive works have been done on CRNs and TDCS-based multiuser access system, but mostly limited to theoretical analysis and simulation. In this paper, a practical testbed setup is designed with two National Instruments (NI) devices (PXIe-1085 and PXIe-1082) with four USRP-RIO devices. The system performance of TDCS-based multiuser access system is evaluated at 2.4 GHz under realistic channel conditions. Specifically, a successive interference cancelation (SIC) receiver is designed to improve its multiuser performance. Two typical scenarios are demonstrated and conducted, i.e., a centralized and a distributed layout. Finally, the bit error ratio (BER) performance of the practical testbed is evaluated and compared with the theoretical results.

The rest of the paper is organized as follows: Sect. 2 introduces the TDCS-based multiuser access system. In Sect. 3, the system setup model, testbed hardware, and SIC receiver are discussed in detail. Numerical simulations and experimental results are presented in Sect. 4. Finally, we summary this paper in Sect. 5.

2 Overview of TDCS Multiple Access

For TDCS-based multiple access, the entire spectrum band is first divided into N spectrum bins according to the spectrum measurement result. Then, a spectrum marking vector $\mathbf{A} = [A_0, A_1 \cdots A_k, \cdots A_N]$ is used to denote the status of spectrum bins. Specifically, the subcarrier A_k will marked as 0 (or 1) if the spectrum measurement result of the k-th spectrum bin is larger (or smaller) than the noise level (threshold).

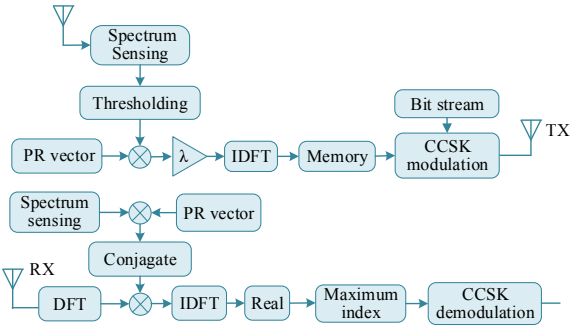


Fig. 1. TDCS transmitter and receiver

To achieve multiple access ability, a user-specific pseudorandom (PR) sequence $\mathbf{P}_u = [e^{j\theta_{u,0}}, e^{j\theta_{u,1}}, \cdots, e^{j\theta_{u,N-1}}]$ is generated for *u-th* user, as shown in Fig. 1. Consider a TDCS with U users, the user-specific PR sequence set are denoted as: $\mathbf{P} = \{\mathbf{P}_1, \mathbf{P}_2, \cdots, \mathbf{P}_u, \cdots, \mathbf{P}_U\}$.

A spectral vector for *u-th* user is obtained by applying element-by-element multiplication operation, ie. $\mathbf{B}_u = \mathbf{A} \cdot \mathbf{P}_u$. Then, a time domain fundamental modulation waveform (FMW) set are generated by performing inverse fast Fourie (IFFT) on the spectral vector, i.e,

$$\begin{aligned} \mathbf{b} &= \{\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_u, \cdots, \mathbf{b}_U\}, \\ \mathbf{b}_{(u)} &= \mathcal{F}^{-1}\left\{\mathbf{B}_{(u)}\right\}, \end{aligned} \tag{1}$$

where $\mathbf{b}_u = \left\{b_u^0, b_u^1, \cdots, b_u^n, \cdots, b_u^{N-1}\right\}$. After IFFT operation, $\mathbf{b}_u$ is subsequently stored and modulated with CCSK modulator. For *M-ary* CCSK, the transmitted signal $\mathbf{z}_u$ is generated by cyclically shifting $\mathbf{b}_u$ to left with $\tau \in \{0, 1, \cdots, M-1\}$ places [1]:

$$\mathbf{z}_u = \left\{z_u^0, z_u^1, \ldots z_u^n, \ldots z_u^{N-1}\right\} = \langle \mathbf{b}_u \rangle_\tau, \tag{2}$$

where $\langle \cdot \rangle_\tau$ denotes the τ points cyclically shift to the left.

At receiver side, after time and frequency synchronization, the received signal $\mathbf{r}$ will be periodically correlated with the local FMW signal $\mathbf{b}$. As the correlator outputs real values, the receiver occupies real operation to deal with the noise. The largest value is selected as the estimate data symbol

$$\hat{\tau} = \arg\max\left\{\Re\left\{\mathcal{F}^{-1}\left\{\mathcal{F}(\mathbf{r}) \cdot \mathcal{F}(\mathbf{B})^*\right\}\right\}\right\}, \tag{3}$$

where $\mathcal{F}$ denotes the FFT operation, $(\cdot)^*$ denotes the conjugate operation and $\Re\{\cdot\}$ is the operation of obtaining real value. For TDCS-based NOMA, the system performance is limited to the multiuser interference. Therefore, by using interference cancelation methods at receiver side can better eliminate the multiuser interference.

3 Testbed Setup and Key Methods

3.1 Testbed Setup and System Model

The testbed setup is implemented with two NI PXIe-1085 and four USRP-RIO-2943 devices, as shown in Fig. 2. PXIe is a PC-based platform designed for measurement. It provides a high-bandwidth backplane and up to 18 hybrid slots to meet the needs of high-performance test and measurement applications. USRP-RIO-2943 supports a center frequency from 1.2 GHz to 6 GHz, with 4 MHz of instantaneous bandwidth. One USRP-RIO-2943 motherboard consists of a field-programmable gate array (FPGA), which delivers a hardware solution for real-time prototyping applications, such as LTE and 802.11 prototyping, Ad-hoc wireless sensor networks and spectrum monitoring.

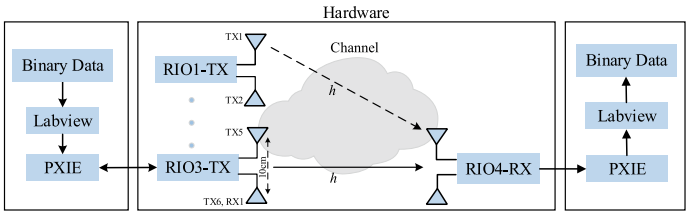


Fig. 2. The main steps from the generation of the data stream to its recovery.

Two identical Zadoff-Chu (ZC) sequences with constant envelope are prepended to the data section for time synchronization. The transmitted streams are split into in phase (I) and quadrature (Q) samples, which are filtered and upsampled before being fed to the RF of the USRP RIO. An upsampler with an upsampling ratio of 4 and pulse shaping filter are used to reduce the intersymbol interference. After upsampling, these samples pass through a root-raised-cosine (RRC) finite impulse response (FIR) filter and then are transmitted through an RF chain.

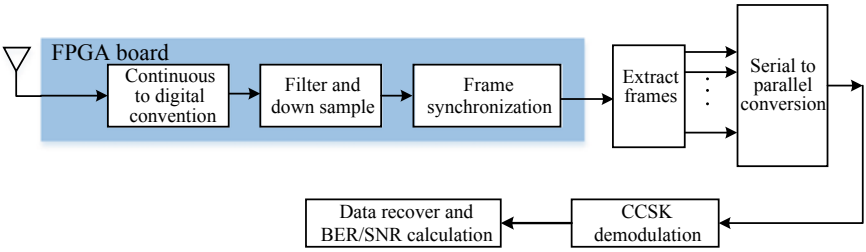


Fig. 3. Time slot design of TDCS-based multiple access system

At receiver side, USRP RIO first records the RF signals and process analog to digital conversion. The signal is downsampled with a sampling ratio of 4 and passes through an RRC-FIR filter, as shown in Fig. 3. Detection and synchronization are implemented before demodulation in TDCS receiver block. The synchronization process can be implemented with FPGA programming of PXIe, which enables real-time processing at the receiver side. Then, the synchronized data are delivered to LabVIEW for further processing.

3.2 Successive Interference Cancelation

For TDCS-based transmission scheme, the transmitted signals are superposed at receiver side and the signal separation of different users is realized via basis function FMWs. The multiuser interference cancelation methods are developed to improve the performance of practical TDCS multiple access system. Successive interference cancelation, parallel interference cancelation, and joint decoding)

are well-known interference cancelation techniques in NOMA scheme [8,9]. joint decoding can achieve the capacity regions for Gaussian multiple access channel and fading multiple access channel. However, SIC has lower complexity than joint decoding, as well as achieves good performance. Thus, a SIC receiver is designed for testbed setup implementation.

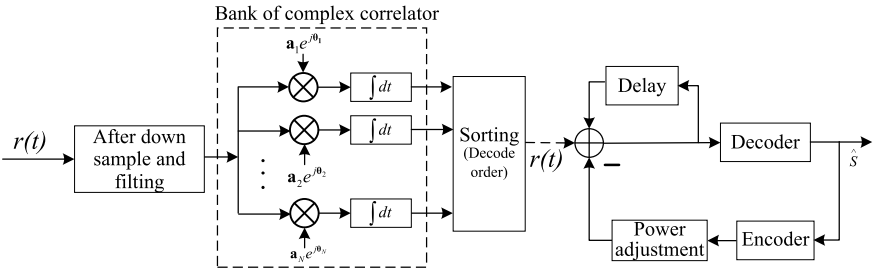
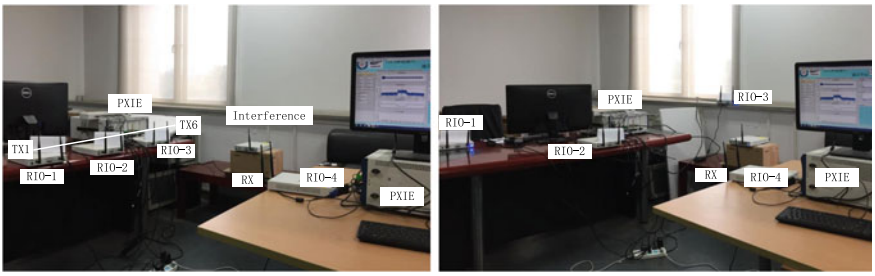


Fig. 4. SIC scheme with CCSK modulation

In SIC, the signal of the first user is decoded by treating the signals of other users as interferences and successively cancel user interferences. The detect order depends on the order of received powers, which are obtained from correlations of the received signal with each users FMWs.

Figure 4 shows the designed SIC receiver for the testbed setup. It is implemented with an interference cancelation module and a user demodulation module, each interference cancelation module is composed of a complex correlator, detector, and signal regenerator (encoder), which can sequentially recover user waveform from the demodulated symbol. The signal regenerator provides a user source estimate for regenerating the transmitted signal in each stage. By suitably selecting the delay, amplitude, and phase, the detected data bits can be remodulated using the corresponding user FMW. Then the remodulated signal is subtracted from $r(t)$. In the multistage configuration, the above process is repeated until all users are fully demodulated.



(a) Centralized layout (b) Distributed layout

Fig. 5. Experimental test environment

4 Experimental Results and Numerical Analysis

4.1 System Measurement

To evaluate the performance of TDCS-based multiple access system with SIC receiver, two typical scenarios are demonstrated and conducted. The system devices are placed in a centralized and distributed layout. All the system tests are carried out in an indoor laboratory. For centralized test, the transmitters are placed directly across the receive antenna, as shown in Fig. 5. Six omnidirectional transmit antennas are placed about 2.1 m away from the receive antenna. All antennas broadcast on a central frequency of 2.4 GHz with system bandwidth 4 Ms/s (data transmission rate). In addition, the transmit and receive antennas are identical and are placed directly across from each other, thus, the channel has a strong line-of-sight (LoS) component.

For distributed test, the transmit antennas will be randomly distributed in the laboratory. At the same time, the distribution of transmit antennas will change at any time. The distributed experimental setup is shown in Fig. 5b.

As shown in Fig. 5, all USRP RIOs are connected to NI PXIe and process the setup synchronization on its FPGA board. The process of data demodulation, BER and SNR calculation are processed in real time. A stream of binary data bits with preamble is sent per transmission frame to evaluate system performance. For SNR calculation, 1280 zeros are added to the end of each frame. To get a more accurate estimation, 1000 frames are evaluated for every BER/SNR point. In addition, the final BER is averaged among all receive antennas and the averaged BER is compared with the theoretical performance. Finally, the performance of SIC is demonstrated via simulations and experiments.

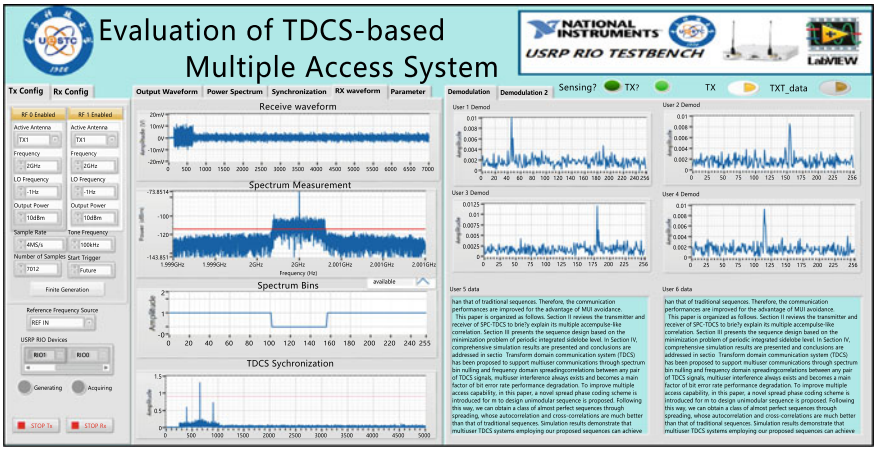
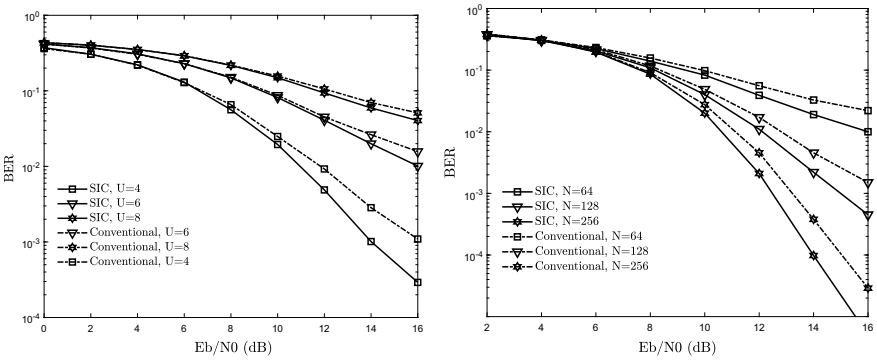


Fig. 6. Displayed system result of TDCS multiuser access with SIC receiver

4.2 Experimental Results

The displayed system result of TDCS multiuser access with SIC receiver is provided in Fig. 6. The displayed waveform including spectrum sensing result, received data waveform, synchronization result, and user demodulation waveform. As mentioned in Sect. 3, two identical ZC sequences with a length of 256 are employed for testbed synchronization. It shows that proposed synchronization method works well with three correlation peaks by employing IFFT/FFT module. The demodulation waveform and recovered message at receiver side are shown on the right side of Fig. 6. The received symbols are demodulated online and in real time.



(a) U=4, 6, 8 (b) N=64, 128, 256

Fig. 7. BER performance comparison of SIC receiver in different situations

Figure 7a shows BER performance of SIC receiver and conventional receiver with sequence length $N = 64$ and $U = 4, 6, 8$. SIC method and conventional single user detection are adopted in the AWGN channel. Simulation result shows that TDCS achieves better performance with SIC receiver than the conventional detection method. Moreover, with the increase of user numbers, the performance of SIC method is closer to conventional user detection. This is caused by the accumulated residual cancelation error. The re-modulated signal is subtracted from the received signal, which may introduce an unknown residual cancelation error.

Figure 7b shows the BER performance of different sequence length when all the spectrum bins are available. The length of sequences ranges from $2^{\wedge 6}$ to $2^{\wedge 8}$ and the number of user $U = 6$. The simulation and experimental results of the testbed setup are demonstrated in Fig. 8. It can be seen that the BER curve of testbed setup experiment result shows a difference from the MATLAB simulation. This can be attributed to unmeasurable noise from USRP and multi-path effect in indoor laboratory. Moreover, due to the near–far effect, the performance

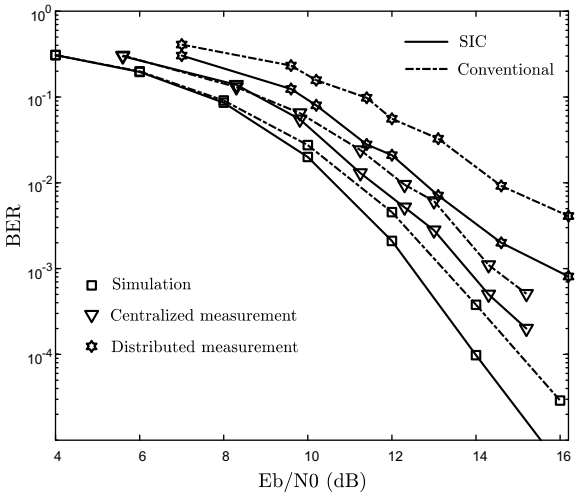


Fig. 8. BER performance of measurement and simulation

of distributed layout of system device is much worse than the centralized layout. The distributed measurement also indicates that the designed SIC receiver effectively suppresses multiuser interference.

5 Conclusion

In this paper, we have presented the prototype design for TDCS-based NOMA system by occupying NI-USRP devices. In particular, the testbed with a SIC receiver has been presented, and system measurement has been described in detail. The designed TDCS with 6-user achieves real-time seamless data transmission and can adaptively adjust its waveform to avoid interference. Moreover, the centralized layout and distributed layout of the experimental setup have been experimentally tested. The simulation shows that a significant gain of about 1 dB to 3 dB can be achieved with proposed SIC receiver. The testbed setup also gives an opportunity to validate more key technologies, such as spectrum sensing, smarter CR frameworks design, and multiuser interference suppression. Future work will also focus on the near–far effect by occupying power control method.

References

1. Hu, S., Bi, G., Guan, Y.L., Li, S.: TDCS-based cognitive radio networks with multiuser interference avoidance. IEEE Trans. Commun. **61**(12), 4828–4835 (2013)
2. Swackhammer, P.J., Temple, M.A., Raines, R.A.: Performance simulation of a transform domain communication system for multiple access applications. In: Military Communications Conference Proceedings, 1999, Milcom, vol. 2, pp. 1055–1059 (2013)

3. Hu, S., Guan, Y.L., Bi, G., Li, S.: Cluster-based transform domain communication systems for high spectrum efficiency. IET Commun. **6**(16), 2734–2739 (2012)
4. Sun, H., Cao, F., Qin, H.: Multiple access applications of transform domain communication system based on phase coding. In: IEEE Fifth International Conference on Big Data and Cloud Computing, pp. 217–222 (2015)
5. Park, Y., Ayoko, G.A., Horvth, E., Kurdi, R., Kristof, J., Frost, R.L.: Practical implementation of spatial modulation. IEEE Trans. Veh. Technol. **62**(9), 4511–4523 (2013)
6. Wei, X., Liu, H., Geng, Z., Zheng, K., Rongtao, X., Liu, Y., Chen, P.: Software defined radio implementation of a non-orthogonal multiple access system towards 5G. IEEE Access **4**(99), 9604–9613 (2017)
7. Hu, S.: Practical implementation of multi-user transform domain communication system for control channels in cloud-based cognitive radio networks. IEEE Access, pp. 99 (2018)
8. Andrews, J.G., Meng, T.H.: Optimum power control for successive interference cancellation with imperfect channel estimation. IEEE Trans. Wirel. Commun. **2**(2), 375–383 (2015)
9. Fang, L., Milstein, L.B.: Successive interference cancellation in multicarrier DS/CDMA. IEEE Trans. Commun. **48**(9), 1530–1540 (2000)

Wireless Information and Power Transmission for Two-Directional Traffic in Spatial-Division Multiple Access Networks

Juan Zhao(✉) and Xiaolong Xu

Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing 210003, PR China
juanzhao.njupt@gmail.com,xuxl@njupt.edu.cn

Abstract. Both the power and the information are transmitted wirelessly, respectively, in downlink and uplink traffics for spatial-division multiple access networks. The receiver in uplink is equipped with multiple antennas to support the multiple access spatially.In the first frequency band, the transmitter sends power wirelessly to multiple users at the same time. In the second frequency band, all the users utilize the received power, respectively, to send information wirelessly to the receiver (the same node in the downlink). The objective is to achieve the best energy efficiency over the whole networks while assuring the required quality of service in terms of the minimum throughput over all the unlink traffics. The unknown parameter is the transmit power in downlink. The difficulty is to deal with the hard derivations over the closed-form solution to the complicated problem. The proposed method successfully obtains the optimal solution in terms of the closed form. Numerical simulations verify the good performance achieved by our proposed method.

Keywords: Power transmission · Spatial-division multiple access · Wireless communications networks · Optimal transmit power

1 Introduction

Wireless communications have changed the whole world substantially. For instance, every mobile user can access the wireless networks almost everywhere at any time for shopping online, surfing online, video watching, and chatting. It

This work was supported in part by the Scientific and Technological Support Project (Society) of Jiangsu Province under Grant BE2016776, the National Key Research and Development Program of China under Grant 2018YFB1003702.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 412–419, 2019.
https://doi.org/10.1007/978-981-13-6264-4_50

is frequently encountered in our routine life that the terminal has no more power yet we have urgently things to do online. Hence, the power charging is highly desirable in this situation yet it is usually not convenient to get these facilities [1,2] due to the mobility of the user especially on the high-speed vehicles. Therefore, the wireless power transmission becomes a very hot topic recently in this literature.

There are over one hundred journal papers over the wireless power transmission from IEEE Xplore. For instance, the reference [3] as one of the earliest journal paper in this field proposes a method for both the power and the information transmission in the same system. The receive power splitting factor is optimized therein, where one received signal is used to extract the information while the other part is used to harvest power. The scenario studied in [3] is limited by the one-directional traffic without consideration over the two-directional traffic. In [4], the studied scenario is reduced from the multiple casting networks as in [3] to the point-to-point networks. The training series for the channel estimation in [4] is taken into consideration during optimization of the receive power splitting factor, where the multiple access problem is not addressed. In [5], the wireless power transmission is examined in the cooperative relaying networks. The power splitting factor across all the relays are optimized by using the game theory. Yet, the result from [5] is only applicable to the one-directional (downlink or uplink) traffic, which will become invalid for the two- directional traffic. In [6], the robust power transmission is researched for the multiple users served by one multi-antenna transmitter. The objective is to maximize the minimum of all the signals to the interference plus the noise by adjusting the power splitting factor. Unfortunately, another directional traffic in these networks is not emphasized. The authors in [7] theoretically analyze the outage probability of the power splitting method in wireless power transmission networks. It is concluded from [7] that the power splitting method is always better than the time splitting method in the same networks. Yet, the energy efficiency is not addressed in [7]. In [8], the diversity gain of the max-min criterion is theoretically examined for the relaying multi-pair communications with power transmission. The conclusion therein is that the max-min criterion proved to be optimal in the non-power transmission system becomes not optimal in the power transmission networks, because the use of the power transmission changes the whole cooperative network fundamentally. Yet, the energy efficiency remains unknown in the related study. All the research results as mentioned above do not address the energy efficiency for the power transmission, which is one of the most key performances in wireless communications especially in the next-generation mobile networks.

Fortunately, the energy efficiency is optimized in [9] for the multi-user multi-input multi-output networks with power transmission. Both the transmit power and the receive power splitting factor are jointly optimized by the Lagrangian relaxation method. The resulted solution is expressed in the nearly analytical form instead of the exactly closed form. This result is suitable for the downlink traffic only yet not applicable to the two-directional traffic simultaneously. The similar work as in [9] is also studied in [10]. The authors in [11] study the energy

efficiency of the wireless power transmission with the eavesdropper existence. They propose a method for the power transmission in the first phase and the information communications in the second phase. However, there is only one mobile user without considering the multiple users from this point of view.

In this paper, we investigate the wireless power transmission in downlink traffic and the wireless information communication in uplink traffic. In downlink traffic, the transmitter in downlink is the same node as the receiver in uplink, which is equipped with multiple antennas. Every mobile user is equipped with single antenna as the receiver for power harvest in the downlink traffic yet as the transmitter for information transmission in uplink traffic. The goal in this paper is to derive the optimal transmit strategy in closed form for the best energy efficiency over the whole networks. The proposed strategy can assure the required quality of service (QoS) in terms of the minimum throughput for the wireless information communications over the whole networks. The difficulty comes from the very complicated problem for the energy efficiency maximization, where the energy efficiency involves the power conversion efficiency at the transmitter in downlink from the plant factory.

The rest of this paper is organized as follows. Section 2 presents the system model for both wireless power transmission and wireless information communications. Section 3 first formulates the mathematical problem for the optimizations, then theoretically derives the optimal solution to the established problem. Section 4 numerically simulates the results achieved by our proposed method. Section 5 concludes the whole paper.

2 System Model

The whole networks consist of two-directional traffics, where the wireless power transmission is done in the first phase or frequency band from the transmitter with $M \geq 2$ antennas from multiple users with each having single antenna. In the second phase or frequency band, the wireless information communication is done from all the mobile users simultaneously to the same receiver. Here, the receiver in the second phase is the same node as the transmitter in the first phase. We assume that the channel state information is perfect at both the receiver and the transmitter in both the directional traffics. The synchronization is also assumed perfect.

In the first phase for the power transmission, the received signal at each mobile user is described as

$$y_k = \sqrt{p_t^{DL}}\mathbf{h}_k^T s + n_k, (k = 1, 2, \cdots, K) \tag{1}$$

where p_t^{DL} is the transmit power from all the antennas on the transmitter in the first phase, which is waiting to be optimized. $\mathbf{h}_k$ $(k = 1, 2, \cdots, K)$ is the wireless channel for power transmission for the kth mobile terminal. $K \geq 2$ is the total number of the mobile users. s is the transmit signal with the unit power in the first phase. n_k is the additive white Gaussian noise (AWGN) with the zero mean

and the unit covariance. The power of the received signal at every user in the first phase is given by

$$p_k = |y_k|^2 = y_k^* y_k = (\sqrt{p_t^{DL}}\mathbf{h}_k^T s + n_k)^*(\sqrt{p_t^{DL}}\mathbf{h}_k^T s + n_k), \tag{2}$$

Thus, the statistical power over all the random channel realizations is expressed as $p_k = Mp_t^{DL}$. The transmit power from each mobile user during the second phase is given by $p_t^{UL} = aMp_t^{DL}$, where $a \in (0, 1)$ is the power conversion efficiency at every mobile user.

In the second phase, the signal received at the receiver for the wireless information communication is given by

$$\mathbf{z} = \sqrt{aMp_t^{DL}}[\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_K]\mathbf{x} + \mathbf{v}, \tag{3}$$

where $\mathbf{g}_k$ is the wireless channel vector from kth mobile user to the receiver in the second phase. $\mathbf{v}$ is the AWGN noise with zero mean and unit covariance. We will formulate the optimization problem to determine the transmit power in the first phase p_t^{DL} with the goal of the energy efficiency maximization with the QoS constraint.

3 Problem Formulation and Optimization

In this section, we first derive the analytical expression of the achievable throughput of all the uplink traffics during the two phases, where the first phase is also taken into consideration even no wireless information is transmitted in this phase. Then a novel definition of the energy efficiency function is presented for the following optimizations as the objective function. Correspondingly, the mathematical problem is formulated to optimize the transmit power adaptively with the varying channel. Key feature is that the proposed method always guarantees the satisfied QoS in terms of the predetermined throughput. The optimal solution is finally derived in the closed form, which facilitates the varying value of the transmit power in downlink traffic.

For the wireless information communications, all the mobile users access the networks by the wireless multiple access channel (MAC), where the receiver in the uplink has $M \geq 2$ antennas to enable this MAC access spatially. Hence, the statistical throughput of the whole networks during the two phases is given by

$$R \approx \frac{m_{min}}{4} \log_2 \Big(1 + \frac{2aM^2 p_t^{DL}}{m_{min}}\Big), \tag{4}$$

where $m_{min} \triangleq \min[M, K]$, p_t^{UL} is the transmit power from each mobile user in the second phase the wireless information communication. In the next generation of the wireless communications networks, the key evaluation performance is the energy efficiency which demonstrates the achieved throughput within unit power cost. The novel metric reflects the energy efficiency for the whole networks.

Therefore, the defined energy efficiency function is expressed for both the energy and the information transmission in the two-directional traffic as

$$\xi \triangleq \frac{m_{min}}{4Mp_t^{DL}} a \log_2 \Big(1 + \frac{2ap_t^{DL}}{m_{min}} M^2\Big). \tag{5}$$

It is still a open problem to optimize the transmit scheme for the best trade-off between the novel energy efficiency goal and the conventional achievable throughput for the studied networks. It is frequently encountered in engineering that there are usually required QoS over the whole wireless networks such as the worst bit or symbol error rate level, the maximum transmit power value, and the minimum achievable throughput. In this paper, our goal is to improve the energy efficiency of the whole green networks yet assures the required QoS in terms of the minimum achievable throughput. Therefore, the corresponding mathematical problem is established as follows:

$$\max_{p_t^{DL}} \quad \frac{m_{min}}{4Mp_t^{DL}} a \log_2 \Big(1 + \frac{2ap_t^{DL}}{m_{min}} M^2\Big), \quad \text{s.t. } R \geq r_o. \tag{6}$$

It is observed from the above expression that the constraint function is concave with respect to p_t^{DL}. The objective function is in the ratio form of a linear function over a $\log(\cdot)$, which results in a non-concave or non-convex cost function. Fortunately, it is proved in the convex book [12] that this form is called as quasi-convex, where the local optimum solution is the globally optimum solution. Hence, the convex optimization methods can be employed to attain the optimal solution to this quasi-concave problem. To obtain the closed form of this solution, the Lagrangian multiplier method is applied to this optimization problem. By calculating the first-order derivative of the corresponding Lagrangian function with respect to the unknown parameter p_t^{DL} and setting it to zero, we have,

$$\log_2 \Big(1 + \frac{2aM^2 p_t^{DL}}{m_{min}}\Big) = \ln 2 \frac{8a^2 M}{m_{min} + 2aM^2} \Big(\frac{\lambda}{p_t^{DL}} + \frac{M}{a}\Big). \tag{7}$$

In order to get the closed form, we resort to approximate the throughput expression by neglecting the constant 1 in the $\log(\cdot)$ function which is widely applied in this literature. Hence, the new equation becomes as a Lambert function, which can be easily solved as

$$p_t^{DL} = \frac{\lambda 8a^2 M (\ln 2)^2}{m_{min}(2aM^2 + m_{min})} \left\{ \mathbb{W}\Big(\frac{\lambda 16 a^3 M^3 (\ln 2)^2}{m_{min}(2aM^2 + m_{min})} \exp\Big\{ - \frac{8aM^2 (\ln 2)^2}{2M^2 + m_{min}} \Big\} \Big) \right\}^{-1}, \tag{8}$$

where $\mathbb{W}(\cdot)$ is the lambert function as a inverse function of $f = xe^x$. For instance, there is a equation $xe^x = f$, where x is the unknown parameter to determine. The solution is given by $x = \mathbb{W}(f)$. λ is a real scalar as the lagrangian multiplier, which can vary for different values of the expected energy efficiency as well as the throughput.

To this end, the optimal transmit power from the transmitter in downlink traffic has been successfully obtained in closed form. This solution achieves the

best energy efficiency for any given required throughput of the uplink traffics in the wireless power in downlink and information communications in uplink traffics. The proposed method is applicable to any number of mobile users for the power harvest in downlink and the information communications in uplink. The scenario of the number of antennas on the transmitter in donwlink traffic is quite general, which includes the conventional multi-input multi-output such as $M = 2, 4$ and the large-scale antennas such as $M = 256, 512$. Also, the mentioned downlink traffic of power transmission can be carried out in the uplink. Correspondingly, the mentioned uplink traffic for information communications can be downlink. The key feature is that the two directional traffics are taken into the consideration for the power and the information wireless transmission.

4 Numerical Simulations

In this section, the numerical simulations are presented to compare the performances of the proposed method for the wireless power and information communications in the spatially multiple access networks. All the wireless channel coefficients are distributed with Gaussian statistics with the zero mean and the unit covariance. All the noises are the AWGN with the zero mean and the unit covariance. The channel realizations are set to 10000. The number of the antennas on the transmitter in downlink is varying from $M = 128{-}256$ and 512. The values of the power conversion efficiency is changing from $a = 10\%$ to $a = 50\%$ and $a = 90\%$. There are $K = 2$ mobile users involved in the power harvest during downlink and the same in the information communication in uplink. Every mobile user has single antenna due to the limited space on the mobile user.

Figure 1 shows the energy efficiency at different values of the achievable throughput, where each curve corresponds to a specified value of the power conversion efficiency at each mobile user. It is observed that each drops with the increase of the achievable throughput. This means the energy efficiency and the throughput conflict with each other. With one increase, the other decreases. It is interesting to note that the drop rate of each curve is different from the counterpart of the other curves. With the improvement of the power conversion efficiency a, the drop rate becomes much larger. For any fixed value of the throughput, the highest curve corresponds to $a = 90\%$, yet the lowest curve corresponds to $a = 10\%$. This is because more power is used as a waste in the low regime of a due to the circuit power cost constant.

Figure 2 presents the relationship between the energy efficiency and the throughput at the fixed value of the power conversion efficiency $a = 50\%$. The three curves correspond to three scenarios of the number of antennas on the transmitter in downlink, respectively, $M = 128$, $M = 256$, and $M = 512$. It is seen from Fig. 2 that with the increasing number of antennas M, the energy efficiency deteriorates for any given value of the required throughput. This is because the power conversion efficiency is not perfect at 100%, which indicates that the more additional power is wasted with the increase of the number of the antennas. Yet, it is observed from Fig. 2 that for any given value of the

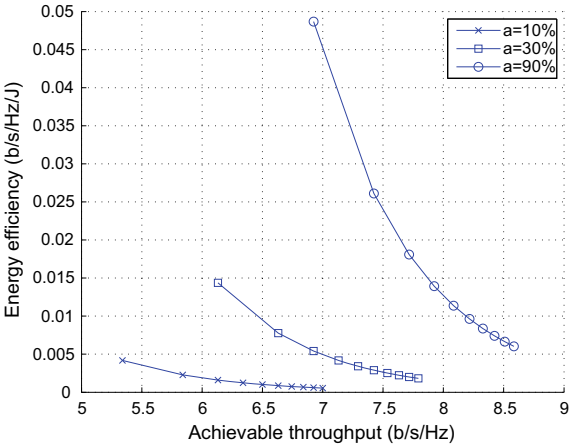


Fig. 1. Energy efficiency versus achievable throughput with the fixed number $M = 128$ of antennas on transmitter in downlink and fixed number of mobile users $K = 2$.

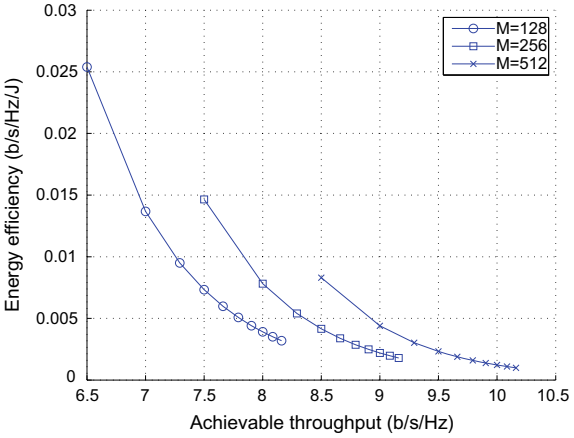


Fig. 2. Energy efficiency as a function of the achievable throughput with the fixed value of the power conversion efficiency at each mobile user $a = 50\%$.

energy efficiency (Y-axis), the corresponding value of the throughput (X-axis) becomes larger with the increase of the number of antennas M. This is because the multiple antennas can support the spatial multiplexing gain.

5 Conclusions

In this paper, we have developed a method for the transmit power optimization in the wireless power transmission in downlink traffic and the wireless information communications in uplink traffic for the spatially multiple access mobile users.

The achievable throughput of the whole networks is expressed statistically over the random channel fading. The energy efficiency function is presented to evaluate the number of the bits within one second and one Hertz under unit power exhaust. The optimal transmit power is successfully obtained in the closed form, where one real scalar can be changed manually to adaptively trade off the energy efficiency and the achievable throughput. The proposed method achieves the best energy efficiency under any given constraint over the required throughput, which enables any number of mobile users to access the whole networks for information communications in uplink.

References

1. Liu, X., Chen, K., Yan, J., Na, Z.: Optimal energy harvesting-based weighed cooperative spectrum sensing in cognitive radio network. Mob. Netw. Appl. **21**, 908–919 (2016)
2. Liu, X., Jia, M., Yan, J., Gao, N., Zhou, J., Lu, W.: Optimal simultaneous cooperative spectrum sensing and wireless power transfer with power splitting in multichannel cognitive radio network. Trans. Emerg. Telecommun. Technol. **26**, 1337–1346 (2015)
3. Khandaker, M.R.A., Wong, K.K.: SWIPT in MISO multicasting systems. IEEE Wirel. Commun. Lett. **3**, 277–280 (2014)
4. Zhou, X.: Training-based SWIPT: optimal power splitting at the receiver. IEEE Trans. Veh. Technol. **64**, 4377–4382 (2015)
5. Chen, H., Li, Y., Jiang, Y., Ma, Y., Vucetic, B.: Distributed power splitting for SWIPT in relay interference channels using game theory. IEEE Trans. Wirel. Commun. **14**, 410–420 (2015)
6. Wang, F., Peng, T., Huang, Y., Wang, X.: Robust transceiver optimization for power-splitting based downlink MISO SWIPT systems. IEEE Signal Process. Lett. **22**, 1492–1496 (2015)
7. Zhou, X., Guo, J., Durrani, S., Krikidis, I.: Performance of maximum ratio transmission in ad hoc networks With SWIPT. IEEE Wirel. Commun. Lett. **4**, 529–532 (2015)
8. Ding, Z., Poor, H.V.: Multi-user SWIPT cooperative networks: is the MaxCMin criterion still diversity-optimal. IEEE Trans. Wirel. Commun. **15**, 553–567 (2016)
9. Shi, Q., Peng, C., Xu, W., Hong, M., Cai, Y.: Energy efficiency optimization for MISO SWIPT systems with zero-forcing beamforming. IEEE Trans. Signal Process. **64**, 842–854 (2016)
10. Vu, Q.D., Tran, L.N., Farrell, R., Hong, E.K.: An efficiency maximization design for SWIPT. IEEE Signal Process. Lett. **22**, 2189–2193 (2015)
11. Yang, W., Mou, W., Xu, X., Yang, W., Cai, Y.: Energy efficiency analysis and enhancement for secure transmission in SWIPT systems exploiting full duplex techniques. IET Commun. **10**, 1712–1720 (2016)
12. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, New York (2004)

Novel Combination Policy for Diffusion Adaptive Networks

Qiang Fan[1], Wang Luo[1(✉)], Wenzhen Li[2], Gaofeng Zhao[1], Qiwei Peng[1], Xiaolong Hao[1], Peng Wang[1], Zhiguo Li[1], Qilei Zhong[1], Min Feng[1], Lei Yu[1], Tingliang Yan[1], Shaowei Liu[1], Yuan Xia[1], Bin Han[1], Qibin Dai[1], Jie Wang[3(✉)], and Guan Gui[3]

[1] NARI Group Corporation/State Grid Electric Power Research Institute, Nanjing, China
luowang@sgepri.sgcc.com.cn
[2] East Inner Mongolia Electric Power Company Limited, Hohhot, China
[3] Nanjing University of Posts and Telecommunications, Nanjing, China
l017010407@njupt.edu.cn

Abstract. Diffusion adaptive networks have received attractive applications in various fields such as wireless communications. Selections of combination policies greatly influence the performance of diffusion adaptive networks. Many diffusion combination policies have been developed for the diffusion adaptive networks. However, these methods are focused either on steady-state mean square performance or on convergence speed. This paper proposes an effective combination policy named as relative-deviation combination policy, which uses the Euclidean norm of instantaneous deviation between intermediate estimation vector of alone agent and the fused estimation weight to determine the combination weights of each neighbor. Computer simulations verify that the proposed combination policy outperforms the existing combination rules either in steady-state error or in convergence rate under various signal-to-noise ratio (SNR) environments.

Keywords: Combination policy · Diffusion adaptive networks · ATC · CTA · Relative-deviation combination policy

1 Introduction

Diffusion adaptive networks consist of a collection of nodes with ability of adaptation, learning, and information exchange. Adaptive networks have many applications ranging from research in biological networks [1, 2] to environmental monitoring and smart city [3]. There are many strategies for these problems in adaptive networks, such as consensus [4–6], incremental [7, 8], and diffusion strategies [9–13]. Compared to incremental strategies and consensus strategies, diffusion strategies are more stable, robust, and scalable [14]. Based on the above reasons, this paper will focus on diffusion implements.

The implementation of diffusion strategies has two phases. One stage named combination stage is that each sink collects and fuses information collected from its

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 420–427, 2019.
https://doi.org/10.1007/978-981-13-6264-4_51

neighbors. Another stage named adaptation stage is that every node in the network updates its own estimation via least mean square algorithms. According to the implementation order of the two stages, two different diffusion algorithms named adapt-then-combine (ATC) diffusion least mean square (LMS) algorithms and combine-then-adapt (CTA) diffusion LMS algorithm have been described in [10]. Whether in ATC diffusion LMS algorithm or in CTA diffusion LMS algorithm, combination weights of every neighbor plays an important role in the performance of adaptive networks.

Many combination policies have been proposed in previous research, such as averaging rule [15], Laplacian rule [16–18], Metropolis rule [16], relative-degree rule [19], relative-variance combination rule [20], etc. These combination rules either have high steady-state error or have low convergence rate. Inspired by the previous construction of combination policy, this paper proposes a new and more effective combination rule. The proposed combination rule bases on Euclidean norm of instantaneous deviation between intermediate estimation vector of single agent and the fused estimation weight to determine combination matrix. Computer simulation results demonstrate that the new combination policy outperforms the existing combination rules in both mean square steady-state error and convergence rate.

Notation: $\mathbb{E}(\cdot)$ denotes the statistic expectation of a vector. $(\cdot)^{\mathrm{T}}$ denotes matrix transpose. $(\cdot)^{-1}$ denotes matrix inverse. The operator $\|\cdot\|$ denotes the ℓ_2-norm of a vector. The operator $|.|$ denotes absolute value function. And, normal font letters denote scalars, boldface lowercase letters denote column vectors, and boldface uppercase letters denote matrices.

The rest of this paper is organized as follows. Section 2 introduces the system model and problem formulation. Traditional combination policies are reviewed and a new and more effective combination policy is proposed in Sect. 3. Computer simulations are conducted in Sect. 4. Section 5 concludes the paper.

2 System Model and Problem Formulation

Consider a collection of N nodes distributed over a space region. They are going to estimate a common parameter $\mathbf{w}^o$ collaboratively which is a column vector of size M. Two nodes that can share information with each other are defined as neighbors. The number of neighbors of node k is called the degree of node k denoted by n_k. And, neighbors of node k is denoted by $\mathcal{N}_k$. At every time instant i, each agent k senses a column signal vector $\mathbf{u}_{k,i}$ and obtains a scalar measurement $d_k(i)$. The covariance matrix of $\mathbf{u}_{k,i}$ is defined as $\mathbf{R}_{u,k} = \mathbb{E}\left\{\mathbf{u}_{k,i}\mathbf{u}_{k,i}^T\right\}$. The relation between the known information $\left\{d_k(i), \mathbf{u}_{k,i}\right\}$ and the unknown parameter $\mathbf{w}^o$ is constructed as follows:

$$d_k(i) = \mathbf{u}_{k,i}^T\mathbf{w}^o + v_k(i), \tag{1}$$

where $v_k(i)$ is additive white Gaussian noise (AWGN) in the environments. $v_k(i)$ is a stationary sequence of independent zero-mean random variables with a finite variance $\sigma_{v,k}^2$.

Each node in the network estimates the parameter $\mathbf{w}^o$ by minimizing local cost function defined as follows:

$$J_k(\mathbf{w}_k) = \mathbb{E}\left|d_k(i) - \mathbf{u}_{k,i}^T \mathbf{w}_k\right|^2. \tag{2}$$

The network seeks the optimal estimation of $\mathbf{w}^o$ cooperatively by minimizing the global cost function which is constructed as follows:

$$J^{global}(w) = \sum_{k=1}^{N} \mathbb{E}\left|d_k(i) - \mathbf{u}_{k,i}^T \mathbf{w}\right|^2. \tag{3}$$

One diffusion strategy for solving above problem adaptively in a distributed manner is ATC diffusion LMS algorithm [21]. The implementation of the diffusion strategy is described in detail below. First, we should design an N × N matrix **A** with nonnegative entries $\{a_{\ell,k}\}$ satisfying the following:

$$\mathbb{1}_{N\times 1}^{\mathrm{T}} \mathbf{A} = \mathbb{1}_{N\times 1}^{T}, \text{ and } a_{\ell,k} = 0 \text{ if } \ell \notin \mathcal{N}_k \tag{4}$$

Here, **A** is combination matrix and $a_{\ell,k}$ denotes the combination weight of node ℓ when it shares information with node k as shown in Fig. 1(a). $\mathbb{1}_{N\times 1}$ is a column vector of size N with all its entries as one. The update equation of ATC diffusion LMS algorithm is defined as follows:

$$\begin{cases} \boldsymbol{\psi}_{k,i} = \mathbf{w}_{k,i-1} + \mu_k \mathbf{u}_{k,i}^* [d_k(i) - \mathbf{u}_{k,i}\mathbf{w}_{k,i-1}] \\ \mathbf{w}_{k,i} = \sum\limits_{\ell\in\mathcal{N}_k} a_{\ell,k}\boldsymbol{\psi}_{\ell,i} \end{cases}, \tag{5}$$

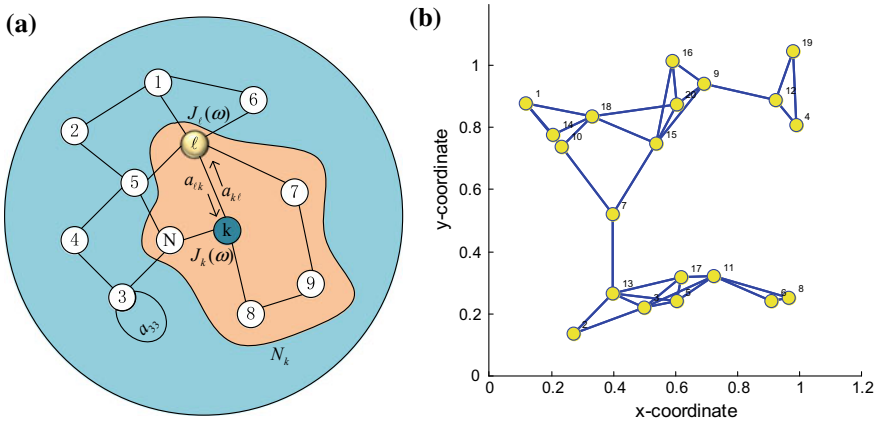


Fig. 1. (a) A diffusion network and the neighborhood of agent k is denoted by the brown field. (b) Topology of the adaptive network.

where μ_k is the step size of node k. The update equation has two steps including incremental update and spatial update as Eq. (5) shows. In incremental update phase, the node k uses its own known data $\{d_k(i), \mathbf{u}_{k,i}\}$ to update its estimated value using stochastic gradient descent method from $\mathbf{w}_{k,i-1}$ to an intermediate estimation $\boldsymbol{\psi}_{k,i}$. In spatial update phase, sink k fuses the intermediate estimation $\{\boldsymbol{\psi}_{k,i}\}$ collected from its neighborhood through combination weights $\{a_{\ell,k}\}$ to obtain the updated weight estimation $\mathbf{w}_{\mathrm{k,i}}$. Naturally, the selection of combination coefficients $\{a_{\ell,k}\}$ influences the performance of the network.

3 Review the Previous Combination Rules and Propose a Novel Combination Policy

Many combination rules have been designed in literature. Three representations of the previous combination policies are described in detail in Table 1. The first two combination policies mainly depend on the degree of the nodes to allocate combination weights. These selections may degrade the performance of adaptive networks because of ignoring the noise profile across a network. Since the nodes accessing to more neighbors may be noisier than the other nodes, it is insufficient to construct the combination weights to their neighbors by only relying on the degree of the nodes. Therefore, it is an important task to design the combination policy by taking into account the noise profile existing in the nodes as well.

Table 1. Three representations of the previous combination policies.

Combination rules	References
$a_{\ell k} = \begin{cases} \frac{1}{n_k}, & \text{if } k \neq l \text{ are neighbors or } k = l \\ 0, & \text{otherwise} \end{cases}$	Uniform policy (15)
$a_{\ell k} = \begin{cases} \frac{n_\ell}{\sum_{m \in N_k} n_m}, & \text{if } k \text{ and } l \text{ are neighbors or } k = l \\ 0, & \text{otherwise} \end{cases}$	Relative-degree policy (19)
$a_{\ell,k}(i) = \begin{cases} \frac{\sigma_{\ell,k}^{-2}(i)}{\sum_{j \in \mathcal{N}_k} \sigma_{j,k}^{-2}(i)}, & \text{if } l \in \mathcal{N}_k \\ 0, & \text{otherwise} \end{cases}$	Relative-variance policy (20)

The third combination rule in Table 1 is named as relative-variance policy which is based on noise quality of each node to determine their combination weights. The noise variance of agents is obtained by relying on instantaneous data approximations and its recursion formula is derived specifically in Table 2. But the relative-variance policy achieves the better performance in terms of steady-state error at a large cost of low convergence rate.

Inspired by the relative-variance combination rule, this paper proposes a more simple and effective combination policy. Instead of the square of Euclidean norm of the instantaneous deviation between $\boldsymbol{\psi}_{k,i}$ and $\mathbf{w}_{k,i-1}$ to estimate noise degree across the

network, we let the Euclidean norm of instantaneous deviation between $\boldsymbol{\psi}_{k,i}$ and $\mathbf{w}_{k,i-1}$ determine the combination weights of each neighbor. The proposed new combination policy named as relative-deviation combination policy is constructed as follows:

$$a_{\ell,k} = \begin{cases} \frac{\left\|\boldsymbol{\psi}_{\ell,i}-\mathbf{w}_{\ell,i-1}\right\|^{-1}}{\sum_{m\in\mathcal{N}_k}\left\|\boldsymbol{\psi}_{m,i}-\mathbf{w}_{m,i-1}\right\|^{-1}}, \textit{for } \ell\in\mathcal{N}_k. \\ 0, \quad \textit{others} \end{cases} \tag{6}$$

In this selection of combination policy, sink k fuses the intermediate information from its neighbors in proportion to the inverses of the Euclidean norm of the instantaneous deviation between $\boldsymbol{\psi}_{k,i}$ and $\mathbf{w}_{k,i-1}$. The combination rule is practically meaningful. $\left\|\boldsymbol{\psi}_{\ell,i}-\mathbf{w}_{\ell,i-1}\right\|^{-1}$ plays the same role as $\left\|\boldsymbol{\psi}_{\ell,i}-\mathbf{w}_{\ell,i-1}\right\|^{-2}$ and is more direct than the latter.

Table 2 The derivation procedure for recursion formula of $\sigma_{\ell,k}^{-2}(i)$

The ATC algorithm is described as follows:
$\boldsymbol{\psi}_{k,i} = \mathbf{w}_{k,i-1} + \mu_k \mathbf{u}_{k,i}^{*}[d_k(i) - \mathbf{u}_{k,i}\mathbf{w}_{k,i-1}]$ (7)
$\mathbf{w}_{\mathrm{k,i}} = \sum_{\ell\in\mathcal{N}_k} a_{\ell,k}\boldsymbol{\psi}_{\ell,i}$ (8)
For sufficiently small step sizes, the estimation $\mathbf{w}_{k,i-1}$ approaches $\mathbf{w}^{o}$, by using model Eq. (1), an equation can be written as follows:
$\boldsymbol{\psi}_{k,i} \approx \mathbf{w}^{o} + \mu\mathbf{u}_{k,i}^{*}v_k(i)$ (9)
And then:
$\mathrm{E}\boldsymbol{\psi}_{k,i} - \mathbf{w}^{o2} \approx \mu^2\sigma_{v,k}^2\mathrm{Tr}(\mathbf{R}_{u,k})$ (10)
Using instantaneous approximation:
$\mathrm{E}\boldsymbol{\psi}_{k,i} - \mathbf{w}^{o2} \approx \boldsymbol{\psi}_{k,i} - \mathbf{w}_{k,i-1}^2$ (11)
Combine (9) and (10) and the recursion formula is derived as follows:
$\sigma_{\ell,k}^2(i) = (1-v_k)\sigma_{\ell,k}^2(i-1) + v_k \times \boldsymbol{\psi}_{\ell,i} - \mathbf{w}_{\ell,i-1}^2$ (12)
where v_k is a positive value close to 1

4 Simulation Results

In this section, we conduct computer simulations to compare the proposed combination policy to previous combination rules including uniform, relative-degree, and relative-variance combination rule. We conduct two experiments in various SNR environments. Among these simulations, the parameter values of the system model and algorithms are uniform except SNR. Parameter values are set as follows (Table 3).

In all experiments, we set uniform step sizes across all agents and all $\mathbf{R}_{u,k}$ equal to each other. The network in this work consists of 20 nodes and the topology of the network is designed following a rule that if the distance of two nodes is less than or equal to a threshold value, the two nodes are defined as neighbors. The topology is generated as shown in Fig. 1(b).

Table 3. Parameter values of the simulations.

Parameters	**Values**
The number of agents	$N = 20$
Input signal	Pseudo-random binary sequences
The length of estimated parameter	$M = 64$
Distribution of estimated parameter	$CN(0, 1)$
Gaussian noise distribution	$CN(0, \delta_n^2)$
Step size of all agents	$\mu = 0.005$

In a real setting, each agent in the adaptive network expending a large geographic region has a distinct separation in the quality of SNR. Taking the practical conditions into consideration, this paper conducts the experiments with each node having different SNR environments. We design two experiments under different noise degree environments to confirm the effectiveness of the new combination policy. Two different noise degree environments are defined as bad noise degree environments (the value of SNR from 5 dB to 10 dB) and general noise degree environments (the value of SNR from 10 dB to 20 dB), respectively. The computer simulation results are shown as follows.

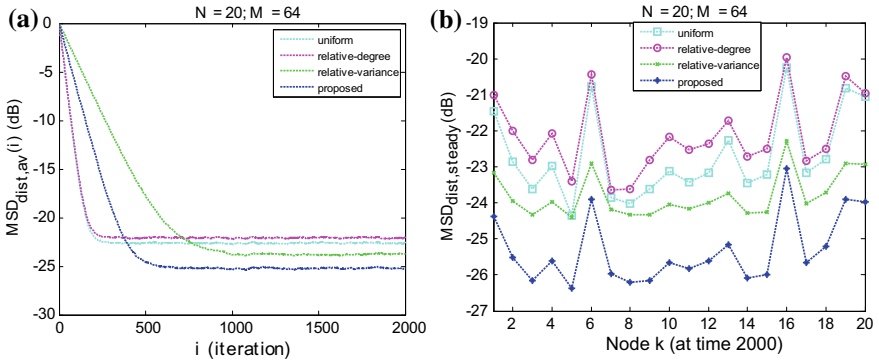


Fig. 2. (a) Average performance of all nodes under different combination policies in bad noise degree environments. (b) Steady performance of every node under bad noise degree environments

Figure 2(a) shows the average performance of all nodes under different combination policies in bad noise degree environments and Fig. 2(b) shows the steady performance of every node under the same conditions. As the results in Fig. 2(a) and (b) demonstrate, whether regarding the average performance of all nodes or the steady performance at each node, relative-variance combination policy outperforms uniform and relative-degree combination rules regarding steady-state error at cost convergence rate. Compared to relative-variance combination policy, our proposed policy named as relative-deviation combination policy obtains superior performance in terms of steady-state error and convergence rate under the bad noise degree environments.

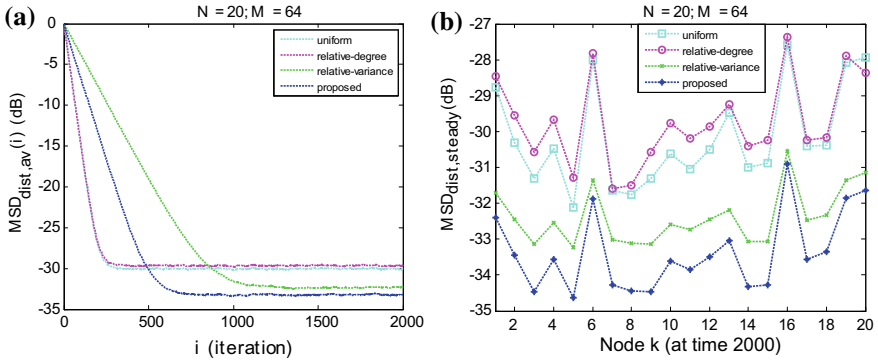


Fig. 3. (a) Average performance of all nodes under general noise degree environments. (b) Steady performance of every node under general noise degree environments

Figure 3(a) shows the average performance of all nodes under different combination policies, in general, noise degree environments and Fig. 3(b) shows the steady performance of every node under the same conditions. As Fig. 3(a) and (b) show, along with the quality of SNR enhancing, the superiority of combination policies based on noise profile across the network including relative-variance and the proposed relative-deviation combination policy significantly improved regarding steady-state error. Compared to relative-variance combination policy, the proposed relative-deviation combination rule keeps considerable merits in both steady-state error and convergence rate under these conditions.

5 Conclusion

This paper proposed an effective combination policy named as relative-deviation combination policy. We conducted two experiments under two different noise degree environments. The computer simulation results demonstrated that the proposed new combination policy outperforms the existing combination rules both in terms of steady-state error and in terms of convergence rate in various SNR environments. The next work is to seek more essential factors in combination policies and to construct more effective combination rules.

Acknowledgements. This research was funded by State Grid Corporation Science and Technology Project (named "Research on intelligent preprocessing and visual perception for transmission and transformation equipment").

References

1. Tu, S.Y., Sayed, A.H.: Mobile adaptive networks. IEEE J. Sel. Top. Signal Process. **5**(4), 649–664 (2011)

2. Cattivelli, F.S., Sayed, A.H.: Modeling bird flight formations using diffusion adaptation. IEEE Trans. Signal Process. **59**(5), 2038–2051 (2011)
3. Chen, L.J., Ho, Y.H., Lee, H.C., Wu, H.C., Liu, H.M.: An open framework for participatory PM2.5 monitoring in smart cities. IEEE Access **5,** 14441–14454 (2017)
4. Nedic, A., Ozdaglar, A.: Distributed subgradient methods for multi-agent optimization. IEEE Trans. Automat. Contr. **54**(1), 48–61 (2009)
5. Kar, S., Moura, J.M.F.: Distributed consensus algorithms in sensor networks with imperfect communication: link failures and channel noise. IEEE Trans. Signal Process. **57**(1), 355–369 (2009)
6. Srivastava, K., Nedic, A.: Distributed asynchronous constrained stochastic optimization. IEEE J. Sel. Top. Signal Process. **5**(4), 772–790 (2011)
7. Rabbat, M.G., Nowak, R.D.: Quantized incremental algorithms for distributed optimization. IEEE J. Sel. Areas Commun. **23**(4), 798–808 (2005)
8. Lopes, C.G., Sayed, A.H.: Incremental adaptive strategies over distributed networks. IEEE Trans. Signal Process. **55**(8), 4064–4077 (2007)
9. Chen, J., Richard, C., Hero, A.O., Sayed, A.H.: Diffusion LMS for multitask problems with overlapping hypothesis subspaces. In: IEEE International Workshop on Machine Learning for Signal Processing, pp. 1–6 (2014)
10. Sayed, A.H.: Adaptive networks. Proc. IEEE **102**(4), 460–497 (2014)
11. Sayed, A.H., Tu, S.Y., Chen, J., Zhao, X., Towfic, Z.: Diffusion strategies for adaptation and learning over networks: an examination of distributed strategies and network behavior. IEEE Signal Process. Mag. **30**(May), 155–171 (2013)
12. Sayed, A.H.: Diffusion adaptation over networks. E-References Signal Process. **61**, 1419–1433 (2013)
13. Chen, J., Sayed, A.H.: Diffusion adaptation strategies for distributed optimization and learning over networks. IEEE Trans. Signal Process. **60**(8), 4289–4305 (2012)
14. Sayed, A.H.: Adaptation, learning, and optimization over networks. Found. Trends Mach. Learn. **7**(4–5), 1–501 (2014)
15. Tu, S., Member, S., Sayed, A.H.: Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks. IEEE Trans. Signal Process. **60**(12), 6217–6234 (2012)
16. Blondel, V.D., Hendrickx, J.M., Olshevsky, A., Tsitsiklis, J.N.: Convergence in multiagent coordination, consensus, and flocking. In: Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference, pp. 2996–3000 (2005)
17. Xiao,L., Boyd, S.: Fast linear iterations for distributed averaging. In: Pmedingr of the 42nd LEEK Conference on UecMon and Conlrd, pp. 65–78 (2003)
18. Scherber, D.S., Papadopoulos, H.C.: Locally constructed algorithms for distributed computations in ad-hoc networks. In: Information Processing in Sensor Networks (IPSN), pp. 11–19 (2004)
19. Xiao, L., Boyd, S., Lall, S.: A scheme for robust distributed sensor fusion based on average consensus. In: Information Processing in Sensor Networks, pp. 63–70 (2005)
20. Cattivelli, F.S., Lopes, C.G., Sayed, A.H.: Diffusion recursive least-squares for distributed estimation over adaptive networks. IEEE Trans. Signal Process. **56**(5), 1865–1877 (2008)
21. Tu, S., Sayed, A.H.: Optimal combination rules for adaptation and learning over networks. In: IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), pp. 317–320 (2011)
22. Cattivelli, F.S., Sayed, A.H.: Diffusion LMS strategies for distributed estimation. IEEE Trans. Signal Process. **58**(3), 1035–1048 (2010)

Link Lifetime and Quality-Based Location Routing for Maritime Wireless Mesh Networks

Jianxiang Zhu, Rongxi He(✉), Cunqian Yu, and Bin Lin

College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China
hrx@dlmu.edu.cn

Abstract. Maritime wireless mesh network (MWMN) has emerged as a cost-effective solution to provide network connectivity for a growing number of maritime users. Routing protocol plays an increasingly important role for the provision of reliable and stable communication in such a network. However, previous works have drawn little attention to the influence of the variable sea surface condition and ship mobility pattern when making routing decisions. In this paper, we propose a novel link lifetime and quality-based location routing (LLQLR) algorithm for an MWMN, with a joint consideration of link lifetime and quality. We first estimate the ship's future location with the help of Kalman filter, then predict the lifetime and quality factor of a link between current ship and its neighbor and calculate the weighted forwarding metric (WFM). Accordingly, the neighbor with maximum value of WFM is selected as the packet forwarder. Finally, we evaluate the performance of our proposed algorithm by extensive simulations and compare it with the existing algorithms. Simulation results show that LLQLR achieves a significant performance improvement with respect to packet delivery ratio and average packet delay variation.

Keyword: Maritime wireless mesh network · Routing · Link lifetime · Link quality · Kalman filter

1 Introduction

Due to the low transmission rate of the traditional maritime communication system, it is hard to satisfy the requirements of maritime users [1]. Recently, the maritime wireless mesh network (MWMN) has emerged to be a cost-effective solution to provide high-speed and economical Internet access. One of the most crucial issues in MWMN is the appropriate decision of a routing protocol, which can contribute to efficient and reliable communication.

Although various routing protocols have been proposed for vehicular ad hoc networks (VANET), few can be directly applied in an MWMN due to the different radio

This work was supported in part by the National Natural Science Foundation of China under Grant 61371091, the "13th five-year" Key Research Project of DMU under Grant 3132016318, and the Fundamental Research Funds for Central Universities under Grant 3132017078.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 428–436, 2019.
https://doi.org/10.1007/978-981-13-6264-4_52

propagations and mobility pattern. In [2], the authors compared the performance of three popular protocols in a maritime environment. The simulation result indicates that ad hoc on-demand multiple distance vector (AOMDV) protocol shows optimal performance. In view of the deficiency of the AOMDV, M-AOMDV protocol was proposed in [3], where multiple routes are maintained in the routing table of each node. Compared with AOMDV, M-AOMDV exhibits better performance in terms of average delay.

In order to exploit the location information obtained through mandatory equipment, several location-based routing protocols were proposed in [4–8]. With a consideration of the path loss model for the maritime environment, the authors in [4] proposed a maritime two-state (MTS) routing protocol. In MTS, ships run in either beaconing state or prediction state depending on the current state of the routing table. Simulation results show that the MTS protocol outperforms M-AOMDV with respect to hop count. With the aim to solve the bandwidth allocation problem, a belief propagation-based cognitive routing strategy was proposed in [5], where collaborative spectrum sensing is achieved by BP algorithm and route selection is done by geographical routing protocols. The proposed scheme shows better performance in terms of average path duration.

As a special case of VANETs, various protocols proposed for VANETs are also applied in MWMNs including greedy perimeter stateless routing (GPSR) [6]. Nonetheless, the established link may be less reliable frequently in GPSR protocol. And, for solving this disadvantage, several proposals have recently emerged [7–9]. The authors in [7] presented an enhancement for the GPSR, named GPSR-R, which exploits information about link reliability to make the routing decision. GPSR + Predict was proposed in [8], which is based on the most promising position-based mechanism. The simulation results show that both GPSR-R and GPSR + Predict scheme can overcome the inherent defects of the GPSR approach. The authors in [9] proposed a strategy named RLBF, where source node predicts the residual link lifetime and chooses the link with the maximum residual lifetime. It can be observed that RLBF exhibits better performance than other traditional strategies.

Unfortunately, the aforementioned routing protocols have drawn little attention to the influence caused by the rough and variable sea surface condition and ship mobility pattern when making routing decisions. Actually, the wave motion may change the antenna height and lead to time-varying received signal quality [4]. Meanwhile, the unique ship mobility pattern will bring about different calculation method for the link lifetime. Therefore, the main objective of this paper is to further improve the GPSR protocol for a reliable and stable communication. In that case, we propose a link lifetime and quality-based location routing (LLQLR) algorithm for an MWMN. It can predict the future location of each ship with the help of Kalman filter [10], and calculate the link lifetime and link quality factor by the future position. With a joint consideration of link lifetime and link quality factor, it calculates the weighted forwarding metric (WFM), and selects the neighbor ship with maximum value of WFM as the suitable forwarding ship. Consequently, the most reliable candidate is obtained, which is beneficial to ensure a stable route between the source and destination.

The remaining of this paper is organized as follows. The system model and routing protocol are described in Sect. 2. Section 3 presents the simulation results, followed by conclusions in Sect. 4.

2 Link Lifetime and Quality-Based Location Routing

2.1 Prediction Mechanism Based on Kalman Filter

Before elaborating the prediction mechanism, we first explain the ship mobility pattern. The authors in [4] presented a general topology of MWMN with a coverage of narrow navigation channels and traffic separation scheme. There are two parallel shipping lanes, one eastbound and the other westbound. The mobility model of ships reflects the real traffic of ships collected from the Strait of Singapore [11]. The same network and mobility model are also applied in our work.

The Kalman filter is an efficient recursive filter to predict the state of a linear dynamic system from a series of noisy measurements [10]. With the help of Kalman filter, we can construct the one-step ahead prediction every Δt time. There exists two crucial vectors in prediction procedure: state vector X_t and measurement vector Z_t. In this paper, the coordinate value of the ship (x_t, y_t) and the magnitude of speed vector v_t form X_t, that is, $X_t = [x_t, y_t, v_t]$. When coming to the next time $t + \Delta t$, based on the aforementioned ship mobility pattern, new state variables can be estimated by

$$\begin{aligned} x_{t+\triangle t} &= x_t \pm v_t \triangle t \\ y_{t+\triangle t} &= y_t \\ v_{t+\triangle t} &= v_t. \end{aligned} \tag{1}$$

For simplicity, we can assume that Δt is equal to 1. Equation (1) means the general form of the process equation, which is provided by [10]

$$X_{t+1} = AX_t + w_t. \tag{2}$$

In addition, we can measure the position of ship based on the data from GPS. Thus, the observation equation is given by [10]

$$Z_t = HX_t + u_t. \tag{3}$$

The transitional matrix A and the measurement matrix H are defined as

$$A = \begin{pmatrix} 1 & 0 & \pm\triangle t \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{4}$$

Let $\hat{X}_{t|t-1}$ and $\hat{X}_{t|t}$ be the a priori estimate and the a posteriori estimate of X_t, respectively. Correspondingly, $P_{t|t-1}$ and $P_{t|t}$ represent the a priori error covariance matrix and a posteriori error covariance matrix. Assume that $\hat{X}_{t|t-1}$ and $P_{t|t-1}$ predicted at the last time are already available, and Z_t is also available by means of the GPS measurement data, then the a posteriori estimate is given by

$$\hat{X}_{t|t} = \hat{X}_{t|t-1} + [P_{t|t-1}H^T(HP_{t/t-1}H^T + R)^{-1}](Z_t - H\hat{X}_{t|t-1}). \tag{5}$$

Then, the state vector matrix at $t + 1$ can be computed as

$$\hat{X}_{t+1|t} = A\hat{X}_{t|t}. \tag{6}$$

Consequently, based on Eqs. (1)–(6), at time t, we can derive the position information at time $t + 1$ in advance.

2.2 Prediction of the Link Lifetime

In general, the node closest to destination is chosen as the forwarder in GPSR protocol. Nonetheless, the selected node may be located at the periphery of the communication range. Considering the dynamic network topology, it will result in unreliable communication. Thus, we define the remaining time that two neighbors are located at the communication range of each other as the link lifetime [9]. And, we develop a link lifetime prediction method in this subsection to promote the routing decision, that is, the ship can predict the link lifetime for time $t + \Delta t$ while the system time is at time t. Due to the unique ship mobility pattern, the calculation method is different from RLBF in [9].

The link lifetime at time $t + \Delta t$ between two ships T and R is decided by relative velocity vector and the positions at time $t + \Delta t$. We assume that $o_T = 1$ if the ship T sails eastward, and $o_T = -1$ if the ship T sails westward. (x_T, y_T) and (x_R, y_R) are the locations of ship T and ship R at time $t + \Delta t$, respectively. Let v_T and v_R represent the speed of ship T and ship R, respectively. With respect to ship T, the magnitude of the two ships' relative velocity vector v_{TR} can be derived as follows:

$$v_{TR} = \sqrt{v_T^2 + v_R^2 - 2v_T v_R o_T o_R}. \tag{7}$$

The direction of the two nodes' relative velocity vector o_{TR} can be given by

$$o_{TR} = \begin{cases} \cos(\arccos\frac{v_T - o_T o_R v_R}{v_{TR}} - \pi) & \text{if } o_T = 1 \\ \cos(\arccos\frac{v_T - o_T o_R v_R}{v_{TR}}) & \text{if } o_T = -1 \end{cases}. \tag{8}$$

Based on geometry theory, we can obtain the calculation equation of link lifetime.

$$LT = \frac{\sqrt{R^2 - d^2\sin^2(\arccos(o_{TR}) - \varphi)} - d\cos(\arccos(o_{TR}) - \varphi)}{v_{TR}}, \tag{9}$$

where R represents the communication range in the recent time and d is the distance between ship T and ship R. Additionally, φ can be computed as

$$\varphi = \mathrm{sgn}(y_R - y_T)\arccos\frac{x_R - x_T}{\sqrt{(x_R - x_T)^2 + (y_R - y_T)^2}}. \tag{10}$$

2.3 Prediction of the Link Quality Factor

Due to the changing sea conditions, the height of an antenna equipped on a ship changes constantly, which results in the instability of the signal reception [4]. In that way, a better routing decision will be available if we can predict the link quality in advance.

Owing to the two-ray interference on the sea surface, the two-ray propagation loss model characterized by the transmitting power P_T and the receiving power P_R can be formulated as follows:

$$P_R/P_T = (\lambda/4\pi d)^2 \times \sin^2(2\pi h_T h_R/\lambda d), \tag{11}$$

where λ and d represent the wavelength and the distance between transmitter and receiver, respectively. Besides, h_T and h_R are the efficient antenna height of the transmitter and receiver, changed with the motion of sea waves.

The motion of sea wave can be approximated by the traveling wave formula [11]. For simplicity, the antenna height of the ship can be computed as the sum of the height of the ship self and the wave height. Suppose that the distance between the ship and a common reference place is d_r. Then, the antenna height h at time t is given by

$$h = \lambda_w/2\pi - H/2 \times \cos(d_r/\lambda - t/T), \tag{12}$$

where H, λ_w, and T represent the significant wave height, average wavelength, and the average wave period, respectively, characterized by the Pierson–Moskowitz [4].

Given the above conditions, if we obtain the location information in advance, we can predict the antenna height of the ship at time $t + \Delta t$ through Eq. (12), then estimate the P_R/P_T at time $t + \Delta t$ through Eq. (11). As P_R/P_T increases, the power loss decreases, and the link quality becomes better correspondingly. Therefore, we define the smaller of P_R/P_T at time t and P_R/P_T at time $t + \Delta t$ as the link quality (LQ) factor, that is,

$$LQ = \min[(P_R/P_T)_t, (P_R/P_T)_{t+\Delta t}]. \tag{13}$$

2.4 Algorithm Description

Since the link lifetime can be affected by the ship mobility pattern, while the link quality factor can be influenced by the motion of sea wave, our proposal will jointly consider the influence of the above two factors. And, we present a weighted forwarding metric (WFM) through the normalization of the above two indicators as follows:

$$WFM = \alpha LT_{norm} + (1 - \alpha) LQ_{norm}, \tag{14}$$

where α represents the weight factor of LT. Each ship broadcasts the hello packets which contain the routing assistance information, such as its current position at time t, the future position at time $t + \Delta t$, the velocity information, the height of the antenna, etc. Thus, based on the routing assistance information from the neighbor ships, ship S can calculate the values of WFM between itself and its neighbors, and then add the values into the corresponding neighbor table. Once a non-destination ship S receives data packets, it will first traverse its neighbor table to confirm whether the destination is in the table. If the neighbor table contains destination, it will directly send the packet to the destination. Otherwise, it will select the ship with the maximum metric value as the forwarder. Algorithm 1 lists the procedure of proposed algorithm.

Algorithm 1. Procedure of proposed algorithm

```
Initially ship S receives a packet, do
    if the packet is hello packet then
        Calculate the LT, LQ and WFM according to Eqs. (12)-(14);
        Add or update the WFM information in the neighbor table;
    else
        if ship S is destination ship then
            Receive the data packet;
        else
            maxWFM ← −∞ and nextHop ← null;
            for each ship N_i in neighbor table of ship S
                if ship N_i is destination ship then
                    ship S forwards the packet to ship N_i;
                else
                    if WFM_i > maxWFM then
                        maxWFM ← WFM_i and nextHop ← N_i;
                    end if
                end if
            end for
            ship S forwards the packet to nextHop;
        end if
    end if
end
```

3 Performance Evaluation

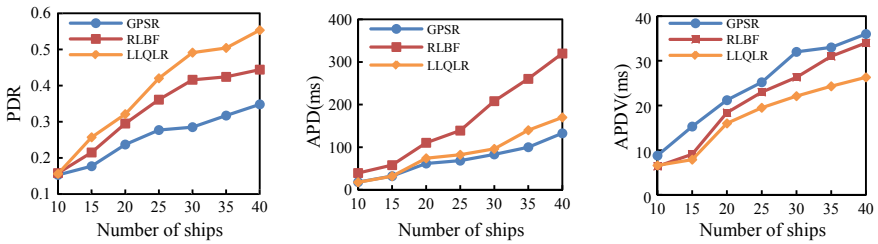
In this section, we evaluate the performance of our proposed algorithm (LLQLR) by extensive simulations based on OPNET modeler 14.5 and compare it with GPSR [6] and RLBF [9] routing algorithms in terms of packet delivery ratio (PDR), average packet delay (APD), and average packet delay variation (APDV).

Both the ship mobility pattern and the path loss model previously presented are used in our simulations. The other simulation parameters are listed in Table 1.

Table 1. Simulation parameters

Parameters	Values	Parameters	Values
Average speed of ships	45 km/h	Antenna model	Omnidirectional
Number of ships	10–40	Traffic type	CBR
Packet size	1024 bytes	Sea state	5
Hello packet interval time	5 s	Simulation time	1200 s

Figure 1(a) shows the performance of the three algorithms in terms of PDR for various numbers of ships. We can observe that the proposed LLQLR outperforms the two other algorithms, followed by RLBF and GPSR in sequence. The main reason is that, without considering the link reliability, GPSR always chooses the ship closest to the destination as a forwarder. Possibly, a link with low residual lifetime has been chosen to forward packets, which may lead to a packet transmission failure. On the contrary, RLBF attempts to select the ship with highest link residual lifetime as a forwarder, which is beneficial to ensure the reliable communication and to reduce the number of packet loss. Accordingly, it has a better PDR performance than GPSR. Unfortunately, none of the above two algorithms considers link quality caused by the variable sea surface conditions that may result in the instability of the signal reception. With a joint consideration of link lifetime and quality in a bi-direction scenario, LLQLR deliberately selects the ship with the highest WFM as a forwarder to ensure a more reliable route than GPSR and RLBF. Potentially, it is helpful to reduce the packet drops, and has the best PDR performance compared with the others.



(a) PDR (b) APD (c) APDV

Fig. 1. The performance of different protocols versus number of ships

Figure 1(b) compares the performance of APD for the three algorithms. The figure shows that GPSR exhibits the best performance, followed by LLQLR and RLBF in sequence. The main reason is that RLBF only focuses on a unidirectional scenario without considering the moving direction of ships. However, in a bi-direction scenario, two ships may move in an opposite direction. If a ship with an opposite moving direction to destination is picked out as the forwarder, more hops will be passed between source and destination. As a result, RLBF has the highest APD. Unlike RLBF, LLQLR improves the calculation of link lifetime for a bi-direction scenario, which

contributes to the reduction of APD. Nevertheless, by choosing a forwarder according to the WFM instead of the distance between the source and destination, LLQLR may bring out an increase in route hop. Consequently, it has a slightly larger value of APD than GPSR.

The performance comparison of APDV is illustrated in Fig. 1(c). With the enhancement of the link stability, there are more chances for a packet to follow the same path. Therefore, the packet endures a comparable delay with a small jitter. Since LLQLR considers the change of sea conditions, and predicts the link quality in advance, it guarantees a more stable route compared with the others and shows the best APDV performance.

4 Conclusion

In order to provide a reliable and stable maritime communication, we propose a novel routing algorithm for an MWMN by taking the influence of the variable sea surface condition and ship mobility pattern into consideration. The proposed algorithm predicts the future location with the help of Kalman filter and calculates the link lifetime and link quality factor between current node and its neighbors. With a joint consideration of link lifetime and link quality factor, the neighbor with maximum value of WFM is chosen as the packet forwarder. Extensive simulation results show that the proposed algorithm achieves a significant improvement in terms of PDR and APDV compared with GPSR and RLBF while with a slightly higher APD value than GPSR.

References

1. Geng, X., Wang, Y., Feng, H., Zhang, L.: Lanepost: lane-based optimal routing protocol for delay-tolerant maritime networks. China Commun. **14**(2), 65–78 (2017)
2. Mohsin, R., Woods, J.: Performance evaluation of MANET routing protocols in a maritime environment. In: 6th Computer Science and Electronic Engineering Conference (CEEC 2014), pp. 1–5. IEEE (2014)
3. Yoo, D., Jin, G., Jang, B., Tuan T., Ro, S.: A modified AOMDV routing protocol for maritime inter-ship communication. In: International Conference on ICT Convergence (ICTC), pp. 605–607. IEEE (2011)
4. Ejaz, W., Manzoor, K., Kim, H.J., Jang, B.T., Jin, G.J., Kim, H.S.: Two-state routing protocol for maritime multi-hop wireless networks. Comput. Electr. Eng. **39**(6), 1854–1866 (2013)
5. Ghafoor, H., Noh, Y., Koo, I.: Belief propagation-based cognitive routing in maritime ad hoc networks. Int. J. Distrib. Sensor Netw. **12**(4), 1–10 (2016)
6. Karp, B.: GPSR: greedy perimeter stateless routing for wireless networks. In: Proceedings of 6th International Conference on Mobile Computing and Networking, pp. 243–254 (2000)
7. Shelly, S., Babu, A.: Link reliability based greedy perimeter stateless routing for vehicular ad hoc networks. Int. J. Vehicular Technol. **2015**, 1–16 (2015)
8. Houssaini, Z.S., Zaimi, I., Oumsis, M., Ouatik, S.E.A.: GPSR + Predict: an enhancement for GPSR to make smart routing decision by anticipating movement of vehicles in VANETs. Adv. Sci. Technol. Eng. Syst. J. **2**(3), 137–146 (2017)

9. Shelly, S., Babu, A.: Link residual lifetime-based next hop selection scheme for vehicular ad hoc networks. EURASIP J. Wirele. Commun. Netw. **2017**(1), 1–23 (2017)
10. Feng, H., Liu, C., Shu, Y., Yang, O.W.: Location prediction of vehicles in VANETs using a Kalman filter. Wirele. Pers. Commun. **80**(2), 543–559 (2015)
11. Wen, S., Kong, P.Y., Shankar, J., Wang, H., Ge, Y., Ang, C.W.: A novel framework to simulate maritime wireless communication networks. In: OCEANS 2007. pp. 1–6. IEEE (2007)

Analysis of End-to-End Communication System Network Model

Kaiyao Zhang(✉), Nan Wu, and Xudong Wang

Information Science and Technology College, Dalian Maritime University, Dalian, China
{zky, wu.nan, wxd}@dlmu.edu.cn

Abstract. Deep learning has been very hot in the field of image recognition and natural language processing in recent years. In this article, we apply deep learning to the field of communication, using neural networks to build end-to-end communication systems under Gaussian channels, and jointly optimize the sender and receiver. The results show that the performance of the end-to-end communication system based on deep learning in Gaussian channel is better than the traditional communication system. Then we change the parameters and results of the network and analyze the results. Finally, we look forward to the next research direction.

Keywords: Autoencoder · Communication system · Deep learning · Physical layer

1 Introduction

The purpose of communication is to transmit messages [1]. In fact, the basic point-to-point communication is to transmit the message of the sender to the receiver through some channel. It is a field of rich expert knowledge included how to design the modulation method, the coding method, how to design the channel model, design the decoding and demodulation methods to achieve the best reception of the signal [3]. The development of the communication field is very mature, and the effect brought about by the improvement of the communication module is not significant, and the optimization of each module alone does not guarantee the overall system is optimal [2], so we use the end-to-end learning system. End-to-end learning system is simply a matter of data processing systems or learning systems. They require multiple stages of processing. End-to-end deep learning ignores all these different phases and replaces them with a single neural network. We use an end-to-end learning model to jointly optimize the transmitter and receiver, allowing the neural network to design modulation and coding method for the current channel environment [4].

In 1943, McCulloch and Pitts abstracted the biological neuron model into the "M-P neuron model" which we have used to date. In recent years, with the growth of computer computing power, the concept of "deep learning" has become popular. "Deep learning" has expanded the neural network to multiple layers, Though adding the neural network brings the generalization ability of the network, but the increase of network parameters also brings the difficulty of network training. Commonly used deep

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 437–444, 2019.
https://doi.org/10.1007/978-981-13-6264-4_53

learning models include Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Generating Confrontation Network (GAN), and Autoencoder (AE) [5]. In this paper, we use the Autoencoder model to model the end-to-end communication system.

The rest of the paper is organized as follows. Section 2 design an end-to-end communication system by autoencoder. In Sect. 3, we changed the structure of the network to explore the impact of network structure on system performance. In Sect. 4, concludes the article.

2 Autoencoder for End-to-End Communication System

One of the simplest communication systems is shown in Fig. 1. Including a transmitter, a channel, and a receiver. The function of the transmitter is to convert various messages into original electrical signals. In order to make the original signal suitable for transmission in the channel, the transmitting device performs some transformation on the original signal, and then sends it to the channel, and the function and transmission of the receiver transmitted. In contrast to the device, it recovers the corresponding original signal from the received signal.

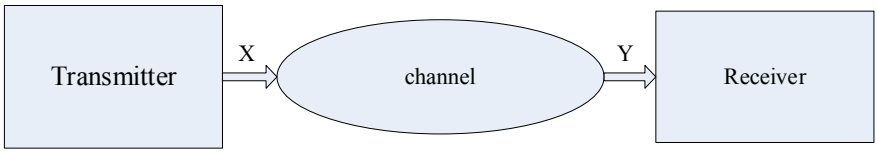


Fig. 1. A simple communications system

For a (n, k) system to send k bit information, the sender wants to send a message s from the message $M = \{1, 2, 3, \ldots, M\}$, $M = 2^k$. The transmitter completes the transformation $s \rightarrow x = f(s)$. The messages are encoded as the one-hot vector. The number of neurons in the last layer of the transmitter is n and the power constraint $\|x\|^2 \leq n$ to ensure that the power of the transmitter is not too large. The channel is a conditional probability distribution $p(y|x)$. Here, our channel model is AWGN $N \sim (0, (2\mathrm{R}\, E_b/N_0)^{-1})$, E_b/N_0 represents the ratio of the signal power E_b per bit to the noise power N_0, R is the communication rate $R = k/n$ [bit/channel use] the task of receiver $y = x + n$ is to reconstruct the message of from the source $y = f(s) \rightarrow \hat{s}$, that is, the last layer of the receiver uses the "softmax" activation function, the output is the probability vector of each neuron, the neuron with the highest probability is set to 1, and the rest is set to 0, and the estimated symbol of the network is $\hat{s}$ obtained. The block error rate (BLER) Pe is then defined as:

$$Pe = \frac{1}{M}\sum_{s} \Pr\left(\hat{s} \neq s|s\right) \tag{1}$$

From the perspective of deep learning, we can think of this end-to-end system model as an autoencoder. The autoencoder is a kind of neural network. After training, you can try to copy the input to the output. There is a hidden layer h in the encoder that can generate coded representation inputs. The network can be seen as two parts: an autoencoder represented by the function $h = f(x)$ and a decoder $\hat{x} = g(h)$ that generates the reconstruction. Here, we modify the structure of the autoencoder slightly. That is, the number of hidden layer neurons in the original autoencoder is less than the number of input layer neurons, because it encodes the original information, and reduces the dimension. In our system, due to the addition of noise, in order to improve the system robustness, we increase the number of hidden layer neurons [6] (Fig. 2).

Figure 3a compares the block error rate (BLER) of a communication system based on an autoencoder to the block error rate with the traditional communication system (with fixed energy constraint $\|x\|_2^2 = n$), where (n, k) is $(7, 4)$ Both systems operate at rate $R = 4/7$, we trained model at a fixed signal-to-noise ratio (SNR) but test at a wide range of SNR. We use Adam as the model optimizer, the learning rate is set to be 0.001, and the batch size is 512. It can be seen that the error block rate of the communication system based on the autoencoder is lower than that of the binary phase-shift keying (BPSK) modulation system without channel coding. And the $(7, 4)$ hard decision system using Hamming code [7].

Figure 3b also performs a comparison similar to Fig. 3a. We set (n, k) to $(2, 2)$, $(6, 6)$, $(8, 8)$, respectively, i.e., $R = 1$, based on the autoencoder over the wide range. The performance of the communication system is better than the unencoded BPSK system. This implies that it has learned some joint coding and modulation scheme, such that a coding gain is achieved. For a truly fair comparison, this result should be compared to a higher order modulation scheme using a channel code (or the optimal sphere packing in eight dimensions). A detailed performance comparison for various channel types and parameters (n, k) with different baselines is out of the scope of this paper and left to future investigations.

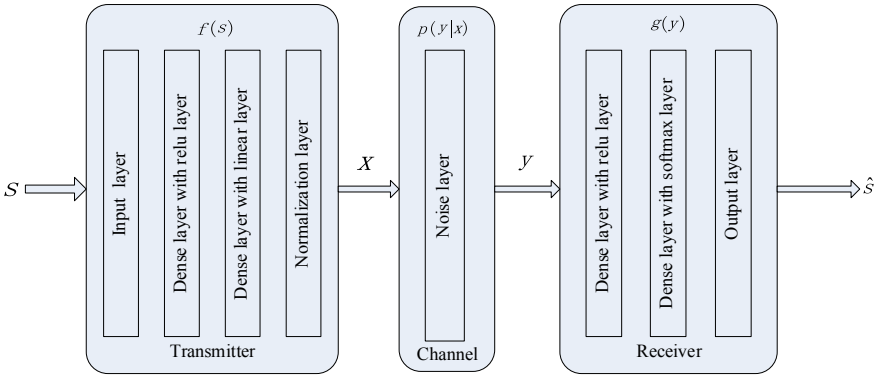


Fig. 2. An end-to-end communication system is regarded as an autoencoder. The channel model is the AWGN channel

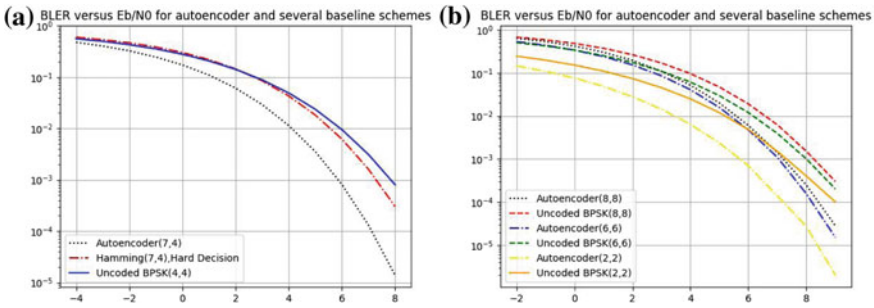


Fig. 3. BLER for the autoencoder and several traditional communication schemes

Here, we regard communication task as a 0,1-classification problem, which converts the information to be transmitted as a one-hot vector. This idea is effective when the amount of information transmitted is small, but when we transmit a lot of bits of information, such as 100 bits, then the one-hot vector space has 2^{100} dimensions, so the network training is extremely difficult [8]. A better approach is to convert the transmitted information into a binary number for transmission, which can reduce the dimension of the vector space and avoid the vector being too sparse. In this case, the Sigmoid activation function is selected as output function, and the binary cross-entropy or MSE is used as the loss function. And this work is in our future exploration.

3 Different Network Structures for Autoencoder

In this part, we changed the network structure [9] to study the impact of changes in network parameters on performance.

3.1 A. Compare Network Performance with Different Activation Function

We adjust the activation function of the original network layer from Relu to Tanh, Sigmoid, respectively, and compare it with the original network. The result is shown in Fig. 4.

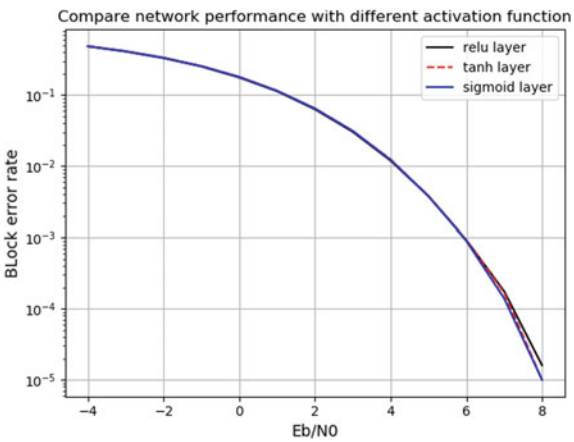


Fig. 4. Compare network performance with different activation function

Figure 4 compares the BLER of the Relu, Tanh, and Sigmoid activation functions. It can be seen from the results that several curves are almost coincident and changing the activation function does not bring about an ideal gain.

Since the Relu function is linear and the gradient is not saturated, the convergence speed will be much faster than Sigmoid/Tanh and the computational complexity is higher than the Sigmoid/Tanh calculation index. Relu only needs a threshold to obtain the activation value. However, it is very fragile during training and may cause neuronal death. For example: Since Relu has a gradient of 0 at $x<0$, this causes a negative gradient to be zeroed at this Relu, and this neuron may never be activated by any data again. If this happens, then the gradient behind this neuron is always 0, that is, Relu neurons are necrotic and no longer respond to any data.

Tanh is a zero mean activation function that can compress data between −1 and 1, but when the input data is too large, the function is apt to saturate, affect the gradient, and make the weights hardly update.

The Sigmoid function can compress the input continuous real value between 0 and 1, but it is easy to saturate. When the input is very large or very small, the gradient of the neuron is close to 0, which makes the reverse in the back-propagation algorithm. Propagation of a gradient close to 0 results in little updating of the final weight. Secondly, the output of Sigmoid is not a zero mean. This will cause the input of the neurons in the posterior layer to be a nonzero mean signal, which will affect the gradient. Assuming that the input of the posterior neurons is positive, then the local gradient is positive for w so that in the process of backward propagation, w is either updated in the positive direction or updated in the negative direction, resulting in a kind of binding. The effect makes convergence slow.

When the independent variable changes within the [−5, 5] interval, the Sigmoid function, and the Tanh function change linearly. When the independent variable changes outside the interval, the function reaches saturation, and the size of the independent variable is changed. The function does not occur greatly. Variety. The threshold of the Relu function is x = 0. When the argument x is greater than 0, the function changes linearly. The outputs of the three network models at the more layers all fluctuate within the [−4, 4] interval. The Sigmoid and Tanh do a linear transformation of the independent variables in the interval and have little effect on the gradient. Therefore, the activation function is changed to Sigmoid or Tanh. The BLER curve does not change significantly.

3.2 B. Compare Network Performance with Different Structure

We added a layer of network layer with the activation function Relu in the transmitter and receiver and compared the performance of the original network structure. The results are shown in Fig. 5.

Figure 5 shows the BLER of the addition of a layer of network at the encoder and an addition layer at the decoder comparing with the structure which is shown in previous section.

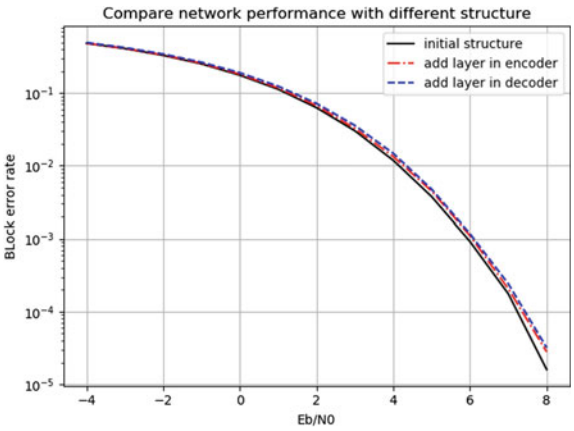


Fig. 5. The BLER of the addition of a layer of network at the encoder and a addition layer at the decoder comparing with the struc-ture which is shown in previous section

At low SNR, the three curves are almost coincident. At high SNR, the original network model performs slightly better. Adding a layer of network at the encoder can learn the characteristics of the input source signal better. Adding a layer of network at the decoder can make the network learn the characteristics of some channels better, but at a high signal-to-noise ratio It can be seen that the signal-to-noise ratio of the original network model is lower. We believe that the source signal is too simple and has fewer features. With a shallow network, we can learn all the characteristics of the signal very well and increase the number of layers of the network. This makes the network complex, and the gradient disappears when it propagates back, or the gradient explodes to affect system performance. Or our test set is not large enough, and it is not precisely with a bit error rate below 10–5.

3.3 C. Compare Network Performance with Different Loss Function

We changed the loss function and set the loss function as the mean square error function and the cross-entropy loss function and compared the BLER curves of the two networks. The results are shown in Fig. 6.

Mean squared error (Mse): $\min C(Y, G(x)) = \|G(x) - Y\|^2 = \sum_i (G(x_i) - y_i)^2$ This loss function used to calculate the Euclidean distance between the predicted value $G(x)$ and the true value Y. The greater the Euclidean distance between the two, the greater the loss, and vice versa. The cross-entropy loss function is used to calculate the difference between the predicted value and the true value.

In the output layer, we use the Softmax nonlinear transformation to calculate the probability of all occurrences. Select the value with the highest probability, set it to 1, and set the rest to 0 as the prediction value. Our real value is the one-hot vector, and the

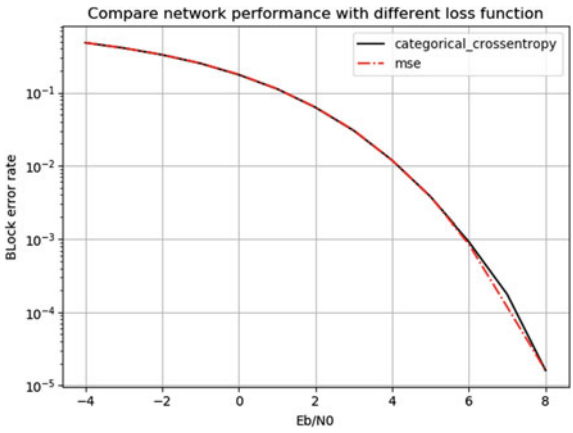


Fig. 6. System performance at mse and categorical crossentropy

corresponding prediction value is also the one-hot vector. Mse is used to calculate the Euclidean distance between the predicted value and the real value. The cross-entropy function is used to calculate the difference between the two. In this experiment, they are similar, so the BLER curve is almost coincident.

4 Conclusion

We validated the idea of end-to-end communication system based on deep learning, showing that the end-to-end system under the Gaussian white noise channel can transmit data without any prior knowledge, and the effect is better than the traditional BPSK system, but how to transmit more bits of information is still a challenge. We believe this is the beginning of extensive research on DL and ML communications, and as the field matures, this may open up possibilities for future wireless communication systems. We have changed some hyperparameters of neural networks to analyze the impact of different parameters on network performance. We believe that the application's survey will not only enable powerful new applications for unsupervised learning in the field of radio communications but also provide insight into the information capacity of similar deep neural networks for many machine learning applications. We can understand the characteristics of the channel by generating a confrontational network. This is our future research.

References

1. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423, 623–656 (1948)
2. Gagliardi, R.M., Karp, S.: Optical communications, 2nd edn. Wiley (1995)
3. Meyr, H., Moeneclaey, M., Fechtel, S.A.: Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing. Wiley (1998)

4. Schenk, T.: RF imperfections in high-rate wireless systems: impact and digital compensation. Springer Science & Business Media (2008)
5. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016)
6. O'Shea, T.J., Karra, K., Clancy, T.C.: Learning to communicate: Channel auto-encoders, domain specific regularizes, and attention. In: IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp. 223–228 (2016)
7. O'Shea, T.J., Hoydis, J.: An introduction to machine learning communications systems. *arXiv preprint* arXiv:1702.00832 (2017)
8. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of AISTATS 2010, vol. 9, pp. 249–256
9. Bengio, Y.: Learning deep architectures for AI. Found. Trends Mach. Learn. **2**(1), 1–127 (2009). (Also published as a book. Now Publishers)

Throughput Analysis for Heterogeneous Networks with Multiple LAAs and Wi-Fi

Yilei Feng[2], Xuesheng Zhou[2], George T. Kweyamba[2], and Zhenzhou Tang[1,2(✉)]

[1] Zhejiang University, Zhejiang, Hangzhou 310027, China
[2] Wenzhou University, Zhejiang, Wenzhou 325035, China
mr.tangzz@gmail.com

Abstract. The Licensed-Assisted Access (LAA) is a novel technology to extend the usable spectrum for Long Term Evolution (LTE). In recent years, there has been dramatic increase in researches on LAA. However, the issue of performance analyses on the heterogeneous network which includes asynchronized multiple LAA networks and Wi-Fi networks remains open. Hence, this paper mainly focuses on the throughput performance of such heterogeneous networks in a saturated scenario.We first model the random backoff procedures of LAA networks and Wi-Fi network into two-dimensional Markov chains. Based on these models, we analyze the throughput performance in the heterogeneous superframe framework which is abstracted from the channel access behaviors of the LAA networks and Wi-Fi networks. The correctness of the analytical results have been verified by sufficient Monte Carlo simulations.

Keywords: LTE in unlicensed spectrum · LTE-U · Licensed-assisted access · LAA · Multi-operator · Performance analyses · Throughput

1 Introduction

With the rapid development of wireless communications, the demand for high-speed and high-quality mobile communication has also rapidly increased. However, applying for new spectrum resources is quite expensive. Therefore, when LTE in unlicensed spectrum (LTE-U) which is a novel technology to extend the LTE into the unlicensed sub-6GHz spectrum was proposed [1], it was welcomed for it is capable of resolving the problem of insufficient spectrum resources.

One of the most concerned issues for LTE-U is the friendly and fair coexistence with Wi-Fi networks since both of them work in the same sub-6GHz spectrum. In order to friendly coexist with Wi-Fi network, the Licensed-Assisted Access (LAA) technology [2] has been proposed and has been included in the 3GPP in Rel-13 [3]. LAA complies with the listen-before-talk (LBT) requirements which is mandated in Europe and Japan.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 445–452, 2019.
https://doi.org/10.1007/978-981-13-6264-4_54

The performance of LAA/Wi-Fi coexistent networks has been extensively studied [4–10]. There have been plenty of works on performance analysis of LTE-U/Wi-Fi coexistence networks in the basis of the Bianchis Markov model [11]. In [4], based on Bianchis Markov model, the throughput of the Wi-Fi Aps and LBT of a cellular BS was investigated. In [5], three-dimensional Markov chain models were exploited to analyze the performance of a coexistent networks composed of LAA and Wi-Fi. In this work, both Cat3 LAA (LAA with fixed backoff window size) and Cat4 LAA (LAA with variable backoff windows size) have been considered. In [6], Bianchis model has been used to calculate the downlink performances under different coexistence scenarios, i.e., LAA/LAA and LAA/Wi-Fi. In [7], the novel heterogeneous superframe analytical framework has been proposed to effectively deal with the asynchronism between the channel access of LAA and Wi-Fi. By jointly using this framework and Bianchi's Markov chain model, the throughput of the LAA/Wi-Fi coexistence network with asynchronous channel access has been accurately derived. Also in [8], an adaptive channel access schemes for the small base stations (SBSs) was proposed and the performance has been evaluated in the basis of the Bianchis Markov model. There are also some work that use stochastic geometry approaches to study the throughput of LAA/Wi-Fi heterogeneous networks. In [9], the stochastic geometry approach was used to evaluate the performance of a large scale deployed LTE-U/Wi-Fi networks. In [10], the performance of a heterogeneous network consisting of an ALOHA-like LTE-U network and a Wi-Fi network has been studied by using stochastic geometry framework.

However, all the works mentioned above only consider the scenario that the coexistence network only contain one LTE-U network. As far as we know that the throughput performance analysis for the heterogeneous network including multiple asynchronous LAA networks remains open. In multi-LAA/Wi-Fi scenarios, the conflict situation will become more complicated because not only the LAA/Wi-Fi interference, but also the conflicts among the different LAA networks shall be considered. In this paper, we first model the random backoff procedures of the LAA networks and the Wi-Fi network into two-dimensional Markov chains. Based on these models, we analyze the throughput performance by utilizing the heterogeneous superframe framework which is abstracted from the channel access behaviors of the LAA networks and Wi-Fi networks. The correctness of the analytical results have been verified by sufficient Monte Carlo simulations.

The rest of the paper is organized as follows. Section 2 introduces the analytical model for the coexistence network in saturated condition. Section 3 analyzes the throughput. Section 4 provides the simulation results and Sect. 5 concludes the paper.

2 Analytical Model

We consider a heterogeneous network which includes two asynchronous LAA networks (denoted as LAA1 and LAA2, respectively) and one Wi-Fi networks. Unless otherwise stated, we use the subscripts l_1, l_2 and w to represent the parameters for the LAA1, the LAA2, and the Wi-Fi network, respectively. Both LAA1 and LAA2 follow the channel access recommendation proposed by 3GPP TR 36.889 [3]. Moreover, we only consider the basic DCF access mechanism regarding the Wi-Fi network. All the channel access behaviors of these three networks can be modeled as Bianchi's Markov chains [11]. Therefore, we can obtain the probability that a node in any of these networks transmits in a randomly selected time slot, denoted as τ is

$$\tau = \frac{(1-p_f^{r+1})^2[(1-2p_f)+q_0 2^m p_f^{m+1}(1-p_f^{r-m})]}{2(1-p_f)^2(1-2p_f)} + \frac{q_0(1-p_f^{r+1})[1-(2p_f)^{m+1}]}{2(1-p_f)(1-2p_f)} \quad (1)$$

where p_f is the probability of a failed transmission caused by collision, r is the maximum number of allowable retransmissions, q_0 is the initial backoff window size and m is the maximum stage of binary exponential backoff.

We adopt the superframe-based analytical framework proposed in [7] to analyze the throughput of the heterogeneous network. Since it is hard to provide a general closed-form solution for LAA/Wi-Fi heterogenous throughput performance with arbitrary timing parameters (mainly the eCCA deferment period T_{eCCA} and the duration of an idle slot σ), we suppose that the $T_{eCCA1} = 22\,\mu s$, $T_{eCCA2} = 25\,\mu s$, $\sigma_{l_1} = 15\,\mu s$ and $\sigma_{l_2} = 12\,\mu s$, respectively. It should be noted that the methodology can be easily extended to the scenarios with other values of timing parameters. The timing relationships among LAA1, LAA2 and Wi-Fi are plotted in Fig. 1.

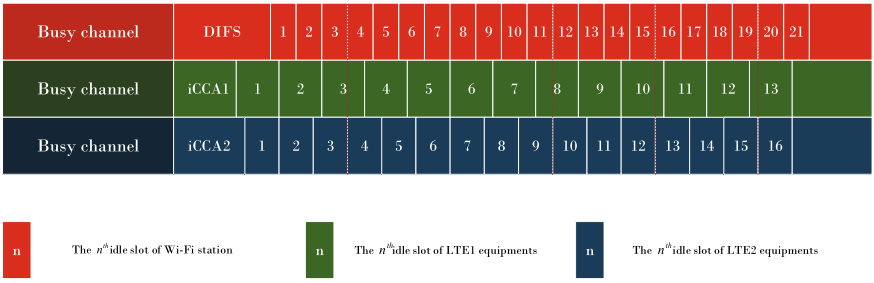


Fig. 1. Timing relation between LAA1 system, LAA2 system and Wi-Fi system after a busy slot.

First, we calculate the probabilities that there are at least n consecutive idle time slots in one superframe for LAA1, LAA2 and Wi-Fi, respectively, denoted as $p_{i,l_1,n}$, $p_{i,l_2,n}$, and $p_{i,w,n}$. The results are shown in Table 1. Thus, p_{f,l_1}, p_{f,l_2},

Table 1. Expressions for $p_{i,l_1,k}$, $p_{i,l_2,k}$ and $p_{i,w,k}$

$I_{l_1} = (1-\tau_{l_1})^{N_{l_1}}$, $I_{l_2} = (1-\tau_{l_2})^{N_{l_2}}$, $I_w = (1-\tau_w)^{N_w}$
$P_{i,l_1,1} = I_{l_1} I_w I_{l_2}$, $P_{i,l_1,n} = P_{i,l_1,n-1} I_{l_1} I_w I_{l_2}^2, n \in \{2\}$
$P_{i,l_1,n} = P_{i,l_1,n-1} I_{l_1} I_w^2 I_{l_2}, n \in \{3,4,7,9,12,13\}$
$P_{i,l_1,n} = P_{i,l_1,n-1} I_{l_1} I_w I_{l_2}, n \in \{5,8,11\}$
$P_{i,l_1,n} = P_{i,l_1,n-1} I_{l_1} I_w^2 I_{l_2}^2, n \in \{6,10\}$
$P_{i,l_1,j+12k} = P_{i,l_2,j} (I_w^{20} I_{l_1}^{12} I_{l_2}^{15})^k, j = 1,2,...,13, k = 0,1,2,...$
$P_{i,l_2,1} = I_{l_2} I_w I_{l_1}$, $P_{i,l_2,n} = P_{i,l_2,n-1} I_{l_2} I_w I_{l_1}, n \in \{2,3,5,8,9,12,14,15\}$
$P_{i,l_2,n} = P_{i,l_2,n-1} I_{l_2} I_w^2 I_{l_1}, n \in \{4,7,10,13\}$, $P_{i,l_2,n} = P_{i,l_2,n-1} I_{l_2} I_w, n \in \{6,11\}$
$P_{i,l_2,n} = P_{i,l_2,n-1} I_{l_2} I_w^2, n \in \{16\}$
$P_{i,l_2,j+15k} = P_{i,l_2,j} (I_w^{20} I_{l_1}^{12} I_{l_2}^{15})^k, j = 1,2,...,16, k = 0,1,2,...$
$P_{i,w,1} = I_w I_{l_1}^2 I_{l_2}^2$, $P_{i,w,n} = P_{i,w,n-1} I_w I_{l_2}, n \in \{2,5,10,12,17,20\}$
$P_{i,w,n} = P_{i,w,n-1} I_w I_{l_1} I_{l_2}, n \in \{4,6,8,9,13,14,16,18,21\}$
$P_{i,w,n} = P_{i,w,n-1} I_w, n \in \{7,15\}$
$P_{i,w,j+20k} = P_{i,w,j} (I_w^{20} I_{l_1}^{12} I_{l_2}^{15})^k, j = 1,2,...,21, k = 0,1,2,...$

and $p_{f,w}$ can be expressed as

$$
\begin{aligned}
p_{f,l_1} &= p_{sf,l_1} \left(\Gamma_{l_1}^{\{2,8,11\}} \Phi_{l_1}^{0,0,1} + \Gamma_{l_1}^{\{1,9\}} \Phi_{l_1}^{0,1,0} + \Gamma_{l_1}^{\{5\}} \Phi_{l_2}^{0,1,1} \right) \\
&\quad + \left(1 - p_{sf,l_1} \Gamma_{l_1}^{\{1,2,5,8,9,11\}} \right) \Phi_{l_1}^{0,0,0} \\
p_{f,l_2} &= p_{sf,l_2} \left(\Gamma_{l_2}^{\{3,9,12,15\}} \Phi_{l_2}^{0,0,1} + \Gamma_{l_2}^{\{1,11\}} \Phi_{l_2}^{1,0,0} + \Gamma_{l_2}^{\{6\}} \Phi_{l_2}^{1,0,1} \right) \\
&\quad + \left(1 - p_{sf,l_2} \Gamma_{l_2}^{\{1,3,6,9,11,12,15\}} \right) \Phi_{l_2}^{0,0,0} \\
p_{f,w} &= p_{sf,w} \left(\Gamma_{w}^{\{2,12,17\}} \Phi_{w}^{1,0,0} + \Gamma_{w}^{\{3,11,15,19\}} \Phi_{w}^{0,1,0} + \Gamma_{w}^{\{7\}} \Phi_{w}^{1,1,0} \right) \\
&\quad + \left(1 - p_{sf,w} \Gamma_{w}^{\{2,3,7,11,12,15,17,19\}} \right) \Phi_{w}^{0,0,0}
\end{aligned}
\tag{2}
$$

where $\Gamma^{\{n_1,n_2,...,n_k\}} \triangleq \sum_{n=1}^{k} p_{i,n}$, $\Phi^{n_1,n_2,n_3} \triangleq [1-(1-\tau)^{N-1} I_{l_1}^{n_1} I_{l_2}^{n_2} I_w^{n_3}]$, p_{sf,l_1}, p_{sf,l_2} and $p_{sf,w}$ are the probabilities of occurring of a superframe in the LAA1, LAA2 and Wi-Fi, respectively.

In the case of saturated traffic, p_{sf,l_1}, p_{sf,l_2} and $p_{sf,w}$ can be obtained as $p_{sf,l_1} = 1/(\mathbf{E}[N_{i,l_1}]+1)$, $p_{sf,l_2} = 1/(\mathbf{E}[N_{i,l_2}]+1)$, and $p_{sf,w} = 1/(\mathbf{E}[N_{i,w}]+1)$ where $\mathbf{E}[N_i]$ is the expectation of the number of consecutive idle slots in one

superframe which can be calculated as follows

$$\begin{aligned}
\mathbf{E}[N_{i,l_1}] &= \sum_{k=1}^{+\infty} k(p_{i,l_1,k} - p_{i,l_1,k+1}) = \frac{\sum_{k=1}^{12} p_{i,l_1,k}}{1 - I_w^{20} I_{l_1}^{12} I_{l_2}^{15}} \\
\mathbf{E}[N_{i,l_2}] &= \sum_{k=1}^{+\infty} k(p_{i,l_2,k} - p_{i,l_2,k+1}) = \frac{\sum_{k=1}^{15} p_{i,l_2,k}}{1 - I_w^{20} I_{l_1}^{12} I_{l_2}^{15}} \\
\mathbf{E}[N_{i,w}] &= \sum_{k=1}^{+\infty} k(p_{i,w,k} - p_{i,w,k+1}) = \frac{\sum_{k=1}^{20} p_{i,w,k}}{1 - I_w^{20} I_{l_1}^{12} I_{l_2}^{15}}
\end{aligned} \tag{3}$$

3 Throughput Evaluation

The throughput can be expressed as the ratio of the average time of successful transmission to the average length of one superframe T_f.

There are two types of busy slots in a superframe. One is a successfully transmitted slot during which only one node is transmitting. We denote the expectation of time used by a success transmission as T_s. And the other is a collision slot during which more than one node is transmitting. We denote the expectation of time consumed by a collision as T_c. A collision slot can be further classified into intra-network collision slot and inter-network collision slot. Thus, $T_f = \mathbf{E}[T_i] + \mathbf{E}[T_s] + \mathbf{E}[T_c]$ and the throughput of LAA1, LAA2, the Wi-Fi and the heterogeneous system are $\mathbf{E}[T_{s,l_1}]/\mathbf{E}[T_f]$, $\mathbf{E}[T_{s,l_2}]/\mathbf{E}[T_f]$, $\mathbf{E}[T_{s,w}]/\mathbf{E}[T_f]$ and $(\mathbf{E}[T_{s,l_1}] + \mathbf{E}[T_{s,l_2}] + \mathbf{E}[T_{s,w}])/\mathbf{E}[T_f]$ respectively.

3.1 Expectation of T_i

First, we calculate $\mathbf{E}[T_i]$. We have obtained $\mathbf{E}[N_i]$ which denotes the expectation of the number of consecutive idle slots in one superframe in (3). So that we can accordingly calculate $\mathbf{E}[T_i]$ as $\mathbf{E}[T_{i,l_1}] = \sigma_{l_1}\mathbf{E}[N_{i,l_1}]$, $\mathbf{E}[T_{i,l_2}] = \sigma_{l_2}\mathbf{E}[N_{i,l_2}]$ and $\mathbf{E}[T_{i,w}] = \sigma_w\mathbf{E}[N_{i,w}]$.

3.2 Expectations of T_c and T_s

We first calculate the probabilities of inter-network collisions. The inter-network collision happens only in the case that multiple nodes belonging to different networks transmit in the time slots which start synchronously. According to the Fig. 1, we have

$$\begin{aligned}
p_{c,(w,l_1)} &= (p_{i,w,2} + p_{i,w,12} + p_{i,w,17})(1 - I_w)(1 - I_{l_1}) \\
p_{c,(w,l_2)} &= (p_{i,w,3} + p_{i,w,11} + p_{i,w,15} + p_{i,w,19})(1 - I_w)(1 - I_{l_2}) \\
p_{c,(l_1,l_2)} &= (p_{i,l_1,1} + p_{i,l_1,9})(1 - I_{l_1})(1 - I_{l_2}) \\
p_{c,(w,l_1,l_2)} &= p_{i,w,7}(1 - I_w)(1 - I_{l_1})(1 - I_{l_2})
\end{aligned} \tag{4}$$

The durations for various inter-network collisions are $t_{c,w,l_1} = \max\{t_{c,w}, t_{c,l_1}\} + t_{def}$, $t_{c,w,l_2} = \max\{t_{c,w}, t_{c,l_2}\} + t_{def}$, $t_{c,l_1,l_2} = \max\{t_{c,l_1}, t_{c,l_2}\} + t_{def}$ and

$t_{c,w,l_1,l_2} = \max\{t_{c,w}, t_{c,l_1}, t_{c,l_2}\} + t_{def}$, respectively, where t_{def} denotes the fixed deferment regulated in the channel access mechanism after the channel state switching from busy to idle, $t_{c,l_1} = \min\{t_{data}, (13/32) \times q \times 10^{-3}\} + \delta$, $t_{c,l_2} = \min\{t_{data}, (13/32) \times q \times 10^{-3}\} + \delta$, $t_{c,w} = t_{data} + \delta$ and δ is the propagation delay. Hence, we can obtain that $\mathbf{E}[T_{c,w,l_1}] = p_{c,(w,l_1)} t_{c,w,l_1}$, $\mathbf{E}[T_{c,w,l_2}] = p_{c,(w,l_2)} t_{c,w,l_2}$, $\mathbf{E}[T_{c,l_1,l_2}] = p_{c,(l_1,l_2)} t_{c,l_1,l_2}$ and $\mathbf{E}[T_{c,w,l_1,l_2}] = p_{c,(w,l_1,l_2)} t_{c,w,l_1,l_2}$.

Next, we calculate the probability of a successful transmission p_s. To this end, the probability that at least one node in the network $\mathrm{x} \in \{l_1, l_2, w\}$ transmits in the n^{th} time slot after the deferment period, denoted as $p_{tr,\mathrm{x},n}$, shall be obtained first. From the Table 1, and Fig. 1, we obtain $p_{tr,l_1,1} = 1 - I_{l_1}$, $p_{tr,l_2,1} = I_{l_1}(1 - I_{l_2})$, $p_{tr,w,1} = I_{l_1} I_{l_2} (1 - I_w)$ and $p_{tr,\mathrm{x},n} = p_{i,\mathrm{x},n-1}(1 - I_{\mathrm{x}}), n = 2, 3, ..., \mathrm{x} \in \{l_1, l_2, w\}$. Therefore, the probabilities of a LAA1 node, a LAA2 node and a Wi-Fi node transmitting in a superframe is

$$
\begin{aligned}
p_{tr,l_1} &= \sum_{k=1}^{+\infty} p_{tr,l_1,k} = (1 - I_{l_1}) \left(1 + \frac{\sum_{k=1}^{12} p_{i,l_1,k}}{1 - I_w^{20} I_{l_1}^{12} I_{l_2}^{15}} \right), \\
p_{tr,l_2} &= \sum_{k=1}^{+\infty} p_{tr,l_2,k} = (1 - I_{l_2}) \left(I_{l_1} + \frac{\sum_{k=1}^{15} p_{i,l_2,k}}{1 - I_w^{20} I_{l_1}^{12} I_{l_2}^{15}} \right), \\
p_{tr,w} &= \sum_{k=1}^{+\infty} p_{tr,w,k} = (1 - I_w) \left(I_{l_2} I_{l_1} + \frac{\sum_{k=1}^{20} p_{i,w,k}}{1 - I_w^{20} I_{l_1}^{12} I_{l_2}^{15}} \right).
\end{aligned}
\tag{5}
$$

Moreover, a successful transmission implies that only one node in the heterogeneous network transmits. Therefore, p_s can be expressed as

$$
\begin{aligned}
p_{s,l_1} &= \frac{N_{l_1} \tau_{l_1} (1 - \tau_{l_1})^{N_{l_1} - 1}}{\left[1 - (1 - \tau_{l_1})^{N_{l_1}} \right]} \left(p_{tr,l_1} - p_{c,(w,l_1)} - p_{c,(l_1,l_2)} - p_{c,(w,l_1,l_2)} \right) \\
p_{s,l_2} &= \frac{N_{l_2} \tau_{l_2} (1 - \tau_{l_2})^{N_{l_2} - 1}}{\left[1 - (1 - \tau_{l_2})^{N_{l_2}} \right]} \left(p_{tr,l_2} - p_{c,(w,l_2)} - p_{c,(l_1,l_2)} - p_{c,(w,l_1,l_2)} \right) \\
p_{s,w} &= \frac{N_w \tau_w (1 - \tau_w)^{N_w - 1}}{\left[1 - (1 - \tau_w)^{N_w} \right]} \left(p_{tr,w} - p_{c,(w,l_1)} - p_{c,(w,l_2)} - p_{c,(w,l_1,l_2)} \right)
\end{aligned}
\tag{6}
$$

The duration of one successful transmission of an LAA node and a Wi-Fi node are $t_{tr,l} = \min\{t_{data}, (13/32) \times q \times 10^{-3}\} + \delta + t_{def}$ and $t_{tr,w} = t_{data} + \delta + SIFS + t_{ack} + \delta + t_{def}$, respectively. Hence, we have $\mathbf{E}[T_{s,\mathrm{x}}] = p_{s,\mathrm{x}} t_{tr,\mathrm{x}}, \mathrm{x} \in \{l_1, l_2, w\}$.

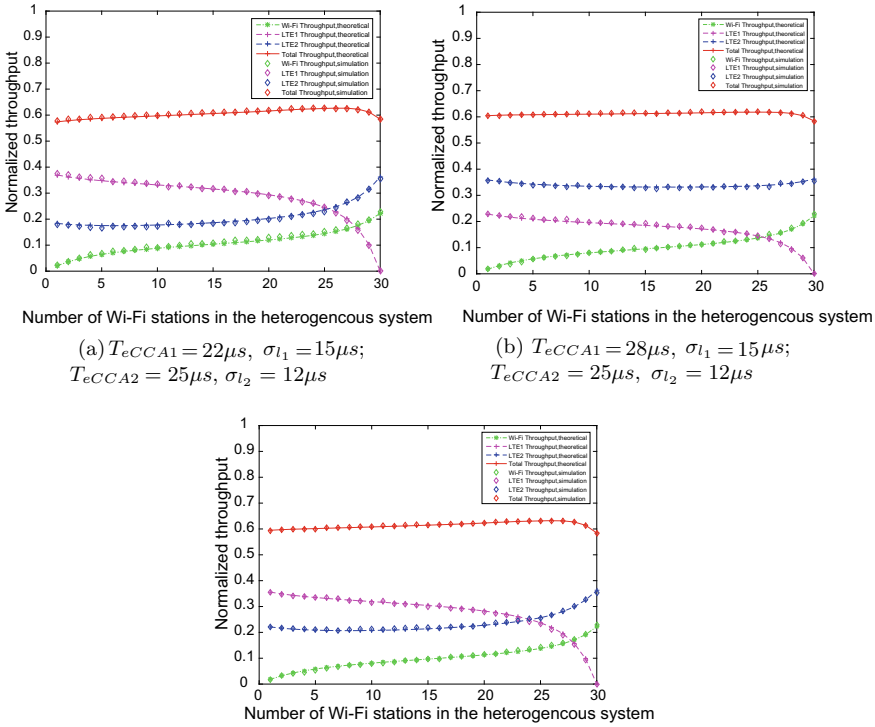
Finally, we consider the intra-network collisions. From the (4) to (6), we have

$$
\begin{aligned}
p_{c,l_1} &= p_{tr,l_1} - p_{c,(w,l_1)} - p_{c,(l_1,l_2)} - p_{c,(w,l_1,l_2)} - p_{s,l_1} \\
p_{c,l_2} &= p_{tr,l_2} - p_{c,(w,l_2)} - p_{c,(l_1,l_2)} - p_{c,(w,l_1,l_2)} - p_{s,l_2} \\
p_{c,w} &= p_{tr,w} - p_{c,(w,l_1)} - p_{c,(w,l_2)} - p_{c,(w,l_1,l_2)} - p_{s,w}
\end{aligned}
\tag{7}
$$

By this far, we can obtain $\mathbf{E}[T_{c,l_1}] = p_{c,l_1} t_{c,l_1}$, $\mathbf{E}[T_{c,l_2}] = p_{c,l_2} t_{c,l_2}$ and $\mathbf{E}[T_{c,w}] = p_{c,w} t_{c,w}$.

4 Numerical Results

Figure 2 shows the throughput of the LAA1 network, LAA2 network, and Wi-Fi network under different values of T_{eCCA} and σ_l. In Fig. 2, there are 40 nodes in the heterogeneous network. The number of nodes in LAA2 is set to a fixed value of 10, and there are altogether 30 nodes in Wi-Fi and LAA1. In this scenario, the frame payload is 8184 bits. The PHY header and MAC header for Wi-Fi a frame is 128 bits and 272 bits, respectively. The length of an ACK frame is 240 bits. The durations for DIFS, SIFS and a Wi-Fi idle slot are 34 µs, 16 µs and 9 µs, respectively. The data rates for LAA and Wi-Fi are both 54 Mbps. The propagation delay (δ) is 1 µs. The maximum number of retransmission (r) is 16. The initial backoff window sizes for LAA and Wi-Fi are 16 and 32, respectively. And the maximum stage of binary exponential backoff is 5. Also in Fig. 2a, $T_{eCCA1} = 22\,\mu s$, $\sigma_{l_1} = 15\,\mu s$, $T_{eCCA2} = 25\,\mu s$ and $\sigma_{l_2} = 12\,\mu s$. In Fig. 2b, $T_{eCCA1} = 28\,\mu s$, $\sigma_{l_1} = 15\,\mu s$, $T_{eCCA2} = 25\,\mu s$ and $\sigma_{l_2} = 12\,\mu s$. And in Fig. 2c, $T_{eCCA1} = 22\,\mu s$, $\sigma_{l_1} = 21\,\mu s$, $T_{eCCA2} = 25\,\mu s$, $\sigma_{l_2} = 12\,\mu s$. The theoretical analytical results and Monte Carlo simulation results are compared. It is shown that



(a) $T_{eCCA1} = 22\mu s,\ \sigma_{l_1} = 15\mu s;$
$T_{eCCA2} = 25\mu s,\ \sigma_{l_2} = 12\mu s$

(b) $T_{eCCA1} = 28\mu s,\ \sigma_{l_1} = 15\mu s;$
$T_{eCCA2} = 25\mu s,\ \sigma_{l_2} = 12\mu s$

(c) $T_{eCCA1} = 22\mu s$, $\sigma_{l_1} = 21\mu s;$
$T_{eCCA2} = 25\mu s,\ \sigma_{l_2} = 12\mu s$

Fig. 2. Normalized throughput, analytical results versus simulations.

the simulation curves and the corresponding theoretical ones coincide with each other pretty well, which illustrates the correctness of our theoretical analysis.

5 Conclusion

In this paper, the throughput performance of a coexistence network consisting of multiple LAA networks belonging to various operators and one Wi-Fi network has been extensively investigated. The backoff procedures of LAA and Wi-Fi have been modeled as 2D Markov chains. The asynchronism in channel access among different networks have been fully considered and the heterogeneous superframe analytical framework has been introduced to solve the problem. Sufficient simulations based on the Monte Carlo method have been performed and the effectiveness of the theoretical analyses have been verified.

Acknowledgement. This work was supported in part by the Natural Science Foundation of China under the Grant 61671329, in part by the Zhejiang Provincial Natural Science Foundation of China under the Grant LY16F010016 and in part by the Fundamental Scientific Research Project of Wenzhou City under Grant G20180008.

References

1. QualComm: Extending LTE advanced to unlicensed spectrum. White paper, December 2013
2. Ericsson, Qualcomm, Huawei, Alcatel-Lucent: Study on Licensed-Assisted Access using LTE. Document 3GPP TSG RAN Meeting #65, RP-141664, September 2014
3. 3GPP: 3GPP TR 36.889: Study on licensed-assisted access to unlicensed spectrum (Release 13). Document, July 2015
4. Song, Y., Sung, K.W., Han, Y.: Coexistence of Wi-Fi and cellular with listen-before-talk in unlicensed spectrum. IEEE Commun. Lett. **20**(1), 161–164 (2016)
5. Gao, Y., Chu, X., Zhang, J.: Performance analysis of LAA and Wi-Fi coexistence in unlicensed spectrum based on Markov chain. In: Global Communications Conference, pp. 1–6 (2017)
6. Chen, C., Ratasuk, R., Ghosh, A.: Downlink performance analysis of LTE and Wi-Fi coexistence in unlicensed bands with a simple listen-before-talk scheme. In: Vehicular Technology Conference, pp. 1–5 (2015)
7. Tang, Z., Zhou, X., Hu, Q., Yu, G.: Throughput analysis of LAA and Wi-Fi coexistence network with asynchronous channel access. IEEE Access **6**, 1–1 (2018)
8. Yin, R., Yu, G., Maaref, A., Li, G.Y.: LBT-based adaptive channel access for LTE-U systems. IEEE Trans. Wirele. Commun. **15**(10), 6585–6597 (2016)
9. Bhorkar, A., Ibars, C., Zong, P.: Performance analysis of LTE and Wi-Fi in unlicensed band using stochastic geometry. In: IEEE International Symposium on Personal, Indoor, and Mobile Radio Communication, pp. 1310–1314 (2014)
10. Wang, X., Quek, T.Q.S., Sheng, M., Li, J.: Throughput and fairness analysis of Wi-Fi and LTE-U in unlicensed band. IEEE J. Sel. Areas Commun. **35**(1), 63–78 (2017)
11. Bianchi, G.: Performance analysis of the IEEE 802.11 distributed coordination function. IEEE J. Sel. Areas Commun. **18**(3), 535–547 (2000)

The Optimized Retransmission Size Choosing of RB-HARQ Implemented in Secure Communication for High Throughput

Lei Wang, Bing Zhao, Daoxing Guo(✉), and Xiaokai Zhang

College of Communications and Engineering,
The Army Engineering University of PLA, Nanjing 210007, China
xyzgfg@163.com

Abstract. A physical-layer coding scheme is designed to enhance the efficiency on the premise of secure communication in this paper. It combines low-density parity-check (LDPC) codes with reliability-based hybrid automatic repeat-request (RB-HARQ) protocol. Specifically, the scheme adapts to the channel conditions and choose the optimized sizes of succeeding retransmissions for high throughput. Moreover, this paper employs the density evolution (DE) method to analyze the message probability density (MPD), which assists in obtaining the approximate bit-error rate (BER) in theory. With the assessment of BER and throughput, simulation results demonstrate the performance of our design.

Keywords: Wire-tap channel · Reliability-based HARQ · Physical-layer security · Low-density parity-check codes · Throughput

1 Introduction

When security is implemented in the physical layer, it can realize the information-theoretically secure communication. In this case, the well-known additive white Gaussian noise (AWGN) wire-tap channel model [1] is considered, where the security is achieved by the differences between the channels experienced by eavesdropper and legitimate receiver. Further study [2] shows a feedback mechanism operates well in the implementation of physical-layer security schemes when Eve's channel quality is better than Bob's. A secure HARQ (S-HARQ) protocol is presented in [3], which analyzes the secrecy and decodability outage probabilities. The study of secure communication [4] utilizes incremental redundancy (INR) secure HARQ protocol with rate adaptation over block-fading wire-tap channels. In [5], a reliability-based HARQ (RB-HARQ) protocol largely increases both the reliability and security in communication with respect to traditional HARQ protocol.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 453–461, 2019.
https://doi.org/10.1007/978-981-13-6264-4_55

However, the existing RB-HARQ protocol only retransmits the same size of unreliable bits, leading to potential insecurity and inefficient use of channel resources. This imposes restrictions on the throughput. The restrictions can be settled by specifying the size and content of successive retransmissions. In [6], the variable-rate HARQ (HARQ-VR) relies on the explicit reduction of the required number of transmitted symbols. As for the literature [7], it shows the throughput achievable with layer-coded HARQ (L-HARQ) based on increasing the coding rate. In this paper, a method for treating the retransmission size as a design parameter is exploited to address the problem of maximizing the throughput.

The proposed scheme combines the average magnitude of a posteriori probability log-likelihood ratios (APP-LLR) of received bits with that of previous retransmitted bits and then finds the optimized size of subsequent retransmissions to get the maximum throughput. Besides, the density evolution (DE) method cooperating with the chase combining is utilized to approximately analyze the maximum throughput in theory. The rest of this paper is organized as follows. System model and performance metrics are presented in Sect. 2. BER analysis and optimizing algorithm are introduced in Sect. 3. The simulation results of the proposed RB-HARQ protocol and its performance analysis are discussed in Sect. 4. Conclusions are drawn in Sect. 5.

2 System Model and Performance Metrics

2.1 Performance Metric

To evaluate the security of communication, we introduce BER over message bits as a measure for security. In [8], an important parameter, the security gap, is considered in the evaluation of wire-tap channels. And it is defined as

$$Sg[dB] = \mathrm{SNR}_{B,\min}[dB] - \mathrm{SNR}_{E,\max}[dB], \tag{1}$$

where $\mathrm{SNR}_{E,\max}$ is the highest signal-to-noise ratio (SNR) to ensure security ($P^E_{\min} \approx 0.5$); $\mathrm{SNR}_{B,\min}$ is the lowest SNR to ensure reliability ($P^B_{\max} \approx 0$). The security gap indicates the minimum difference of two channel to ensure secure and reliable communication. So the security gap is supposed to be efficiently small to ensure secrecy at the physical layer.

But with the help of RB-HARQ, Bob has the advantage of requesting retransmissions. For this reason, another parameter, the signal-to-noise ratio gap SNR_g [5] is advanced, defined as the difference between the signal-to-noise ratios of Bob and Eve

$$SNR_g[dB] = \mathrm{SNR}_B[dB] - \mathrm{SNR}_E[dB], \tag{2}$$

where SNR_g satisfies $SNR_g > S_g$ to ensure security and reliability. With the different SNR_g, the reliable (R) and secure (S) regions are defined as the ranges of SNR where ${P_e}^B \leq P^B_{\max} \approx 0$ and ${P_e}^E \geq P^E_{\min} \approx 0.5$, presented in Fig. 1.

2.2 System Model

In the model depicted in Fig. 2, **u** denotes a k-bits message. It is scrambled by multiplying scrambling matrix S and encoded into a n-bits codeword **c** using the regular LDPC codes [2,5]. Then, the encoded message **c** is transmitted through different AWGN channels, and Bob/Eve receives $\mathbf{c_B}/\mathbf{c_E}$, respectively.

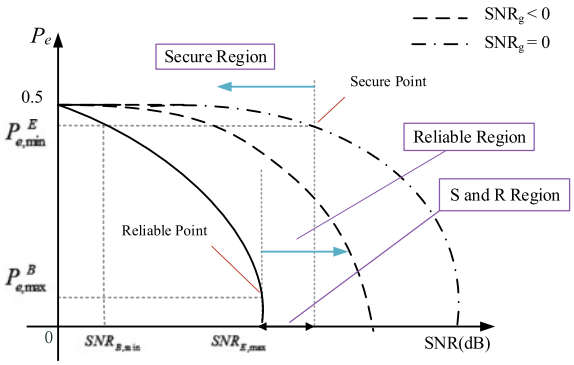


Fig. 1. The secure and reliable region with different SNR_g.

After decoding, Bob and Eve obtain two estimates of the message vector, noted by $\mathbf{u_B}$ and $\mathbf{u_E}$. At last, with multiplication by S^{-1}, the decoded codes can be more likely to go wrong, because descrambling can propagate the residual errors and narrow the security gap [2].

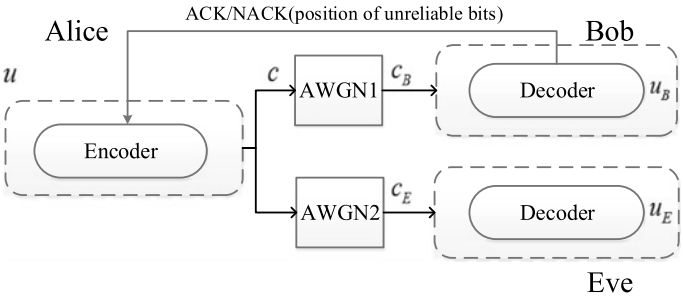


Fig. 2. Gaussian wire-tap channel model with a feedback channel.

When the message $\mathbf{u_B}$ is the same as the source message **u** or the number of retransmission reaches the maximum M, Alice begins to transmit a new round of bits. When not, Bob requests for a retransmission according to its own need. Based on the characteristic of RB-HARQ, the retransmitted bits are the most unreliable bits of a certain size for Bob but may not be the required bits of

Eve. In our study, we assume the ideal feedback channel from Bob to Alice. And a log-likelihood ratio belief propagation (LLR-BP) decoder of LDPC codes is applied, which can output soft-decision APP-LLR values of each bit after iterative decoding process.

Consider the traditional RB-HARQ protocol in physical layer. The sizes of the transmitted bits are fixed so that the scheme faces the inefficient use of the channel. So the sizes can be optimized to realize the goal of maximizing throughput. Then a_m denotes the proportion of the retransmitted bits within the whole packet in the mth retransmission, where $m = 1, 2, \cdots, M$ and M is the maximum.

The user throughput η [9] is defined as the ratio of the average number of successfully received information bits $\bar{N}_b$ to the average number of transmitted channel bits $\bar{N}_s$ after up to M retransmission attempts, given by

$$\eta = \frac{\bar{N}_b}{\bar{N}_s} = \frac{k(1 - P_{f(M)})}{\sum_{m=1}^{M} \bar{N}_{s,k}}, \tag{3}$$

where $\sum_{m=1}^{M} \bar{N}_{s,k} = n + \sum_{m=2}^{M} a_m n P_{f(m-1)}$, $\bar{N}_{s,k}$ is the expected number of codes that are used in the mth retransmission attempt, P_s is the final probability of the packet successful received with the RB-HARQ, and $P_{f(m)}$ is the frame-error rate (FER) in the mth retransmission where $m = 0$ and $m = M$ indicate the first transmission and the final retransmission, respectively.

3 BER Analysis and Optimizing Algorithm

This section demonstrates how to get the maximum throughput under constraints on allowed FER. That requires finding the optimal retransmitted bits' proportion in every retransmission.

For (d_v, d_c)-regular LDPC codes, where d_v and d_c are defined as variable and check node degrees. And in the sum-product algorithm, $P_{i,m}^{[l]}$ and $Q_{j,m}^{[l]}$ represent the message probability density (MPD) of random variables in the lth iteration $v_i^{[l]}$ and $u_j^{[l]}$ after mth retransmission, respectively. $u_j^{[l]}$ and $v_i^{[l]}$ are thought of as the "message" in the log-likelihood ratio which is exchanged between the check nodes and the variable nodes. During the iteration process, L is the maximum number of iterations. When $l = 0$, the variable nodes obtain the initial"message" from the channel. The density of the initial "message" is denoted as $P_{i,0}^{[0]}$ obeying Gaussian distribution with a mean value of $2/\sigma^2$ and a variance value of $4/\sigma^2$.

Based on the above notations, the density evolution (DE) method can be used to perform the analysis of the BER, and the FER can be obtained through the relationship with the BER [10]. According to DE, the density of random variables $v_i^{[l]}$ in the l-th iteration without retransmitted message is given by

$$P_{i,0}^{[l]} = P_{i,0}^{[0]} \otimes \sum_{i \geq 2} v_i [Q_{j,0}^{[l-1]}]^{\otimes (i-1)}, \tag{4}$$

where $\otimes$ indicates convolution and the calculation of $Q^{[l]}$ can be obtained in [10]. $P_{i,m}^{[L]}$ updates after receiving the retransmitted information. When the user receives the retransmitted message, the chase combining strategy used in [11] combines $P_{i,0}^{[L]}$ in the first transmission with $P_{i,m}^{[L]}$ of following retransmissions in the decoding process. For all the variable nodes, the details of calculation are presented as follows.

First, separate the L-iteration MPD $P_{i,m-1}^{[L]}$ in $(m-1)$-th retransmission into two parts which are the MPD $P_{i,m-1(retran)}^{[L]}$ of the most unreliable bits and the density $P_{i,m-1(unretran)}^{[L]}$ of the remaining bits, and then normalize them. The most unreliable bits account for proportion a_j of all the coded bits, which are the required bits in the next retransmission, and accordingly the remaining bits account for proportion $1 - a_j$.

Second, calculate the MPD of the retransmitted bits, which is equal to the convolution of the MPD $P_{i,m-1(retran)}^{[L]}$ of the required bits and the MPD $P_{i,0}^{[0]}$ obtained from AWGN channel.

Third, combine the MPDs of the remaining and the retransmitted in different proportions and the combined MPD is normalized as the initial MPD $P_{i,m}^{[0]}$ in the m-th retransmission, which is approximate Gaussian distribution.

Therefore, the message density of all the variable node after m-th retransmission can be expressed as follows:

$$P_{i,m}^{[0]} = a_j \times (P_{i,0}^{[0]} \otimes P_{i,m-1(retran)}^{[L]}) + (1 - a_j) \times P_{i,m-1(unretran)}^{[L]} \tag{5}$$

Then, we can measure the BER and the FER performance in the m-th retransmission by the following equation:

$$\begin{aligned} \bar{P_{e,m}}|_{B/E} &= \frac{\sum\limits_{i \in A} (\int_{-\infty}^{0} P_{i,m}^{[L]}(v)|_{B/E} dv)}{K} \\ &= \frac{\sum\limits_{i \in A} (\int_{-\infty}^{0} P_{i,m}^{[0]}|_{B/E} \otimes \sum\limits_{i \geq 2} v_i [Q_{j,m}^{[L-1]}(u)]^{\otimes(i-1)} dv)}{K} \end{aligned} \tag{6}$$

$$\bar{P_{f,m}} = 1 - (1 - \bar{P_{e,m}})^k \tag{7}$$

Considering a perfect scrambling, the BER after descrambling equals half the FER expressed by [3]:

$$\bar{P_{s,m}}|_{B/E} = \frac{1}{2} \bar{P_{f,m}}|_{B/E} \tag{8}$$

where $\bar{P_{s,m}}$ denotes the BER after descrambling.

When the maximum of the retransmission times is fixed as M, the optimized $[a_1, a_2, ..., a_M]$ is exclusive. So the Eq. (2) can be formulated as

$$[a_1, a_2, ..., a_M] = \arg \max_{a_1, a_2, ..., a_M} \frac{k[1 - P_{f,M}(a_1, a_2, ..., a_M)]}{n + \sum_{m=2}^{Q} a_m n P_{f,m-1}(a_1, a_2, ..., a_{m-1})}. \tag{9}$$

To ensure the reliability, the BER $\bar{P_{s,m}}|_B$ should satisfy the condition ($\bar{P_{s,m}}|_B \leq \varepsilon_t$), where ε_t should be small enough. So the allowed FER $P_{f,M}|_B$ should meet the requirement $P_{f,M}|_B \leq 2\varepsilon_t$ in the case of throughput maximization under constraints [10], that is

$$\eta_{\varepsilon_t} = \max_{a_1, ... a_M} \eta, \qquad s.t. \qquad P_{f,M}|_B \leq 2\varepsilon_t \tag{10}$$

The solution is expressed as

$$\eta_{\varepsilon_t} = \begin{cases} \eta, & \text{if} \quad P_{f,M} \leq 2\varepsilon_t \\ \frac{1-2\varepsilon_t}{U(2\varepsilon_t)}, & \text{otherwise,} \end{cases} \tag{11}$$

where the first case corresponds to the original solution. Then, the scheme performs to find the maximum throughput by using an exhaustive search to find the global optimized sizes in theory.

4 Simulation Results

In this section, simulation results are presented for comparing the performance of the system implementing the RB-HARQ protocols. In the considered example, a (375, 500) regular LDPC codes is applied, with fixed variable node degree $d_v = 3$ and check node degree $d_c = 12$. The maximum number of retransmissions for both protocols is $M = 4$. And 100 decoding iterations and 10 iterations were used in the simulation and density evolution, respectively. For comparison, the fixed proportions of retransmitted bits for traditional RB-HARQ protocols are $a = 0.1$. And for the proposed protocol, the size of retransmission is only allowable for the following candidates: $a_m \in [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$.

We search the candidates of retransmission proportion and find the optimized proportions to obtain the maximum throughput at different SNR conditions in theory. The maximum throughput in theory is obtained by (3), and $P_{f,m}$ is approximately derived by the density evolution analysis. For simulation, the analysis is realized through the definition of throughput and $P_{f,m}$ equals the ratio of the number of wrong frames to the total. Besides, we take the maximum throughput under constrain into consideration. Because of the reliability requirement $\bar{P_{s,m}}|_B \leq 10^{-3}$, the FER satisfies the constrain $P_{f,M} < 2 \times 10^{-3}$. The signal-to-noise ratio gap satisfies $SNR_g = 0$. The results are shown in the

Table 1. The optimized proportion for maximizing throughput at different SNR conditions.

SNR	a_1	a_2	a_3	a_4
−3.0 dB	0.8/0.9	0.4/0.7	0.2/0.5	0.2/0.5
−2.5 dB	0.8/0.8	0.3/0.5	0.2/0.4	0.2/0.3
−2.0 dB	0.7/0.6	0.4/0.4	0.1/0.3	0.1/0.3
−1.5 dB	0.6/0.6	0.2/0.4	0.1/0.2	0.1/0.2
−1.0 dB	0.4/0.4	0.2/0.2	0.1/0.2	0.1/0.2
−0.5 dB	0.4/0.3	0.1/0.2	0.1/0.1	0.1/0.2
0 dB	0.2/0.3	0.1/0.1	0.1/0.1	0.1/0.1
0.5 dB	0.1/0.1	0.1/0.1	0.1/0.1	0.1/0.2
1.0 dB	0.1/0.1	0.1/0.1	0.1/0.1	0.1/0.1
1.5 dB	0.1/0.1	0.1/0.1	0.1/0.1	0.1/0.1
2.0 dB	0.1/0.1	0.1/0.1	0.1/0.1	0.1/0.1
2.5 dB	0.1/0.1	0.1/0.1	0.1/0.1	0.1/0.1
3.0 dB	0.1/0.1	0.1/0.1	0.1/0.1	0.1/0.1

Table 1. In this table, the former data of the retransmission proportion is from the maximum throughput scheme and the latter is from the scheme under constrains.

Figure 3 displays throughput versus SNR and compares the throughput performance. "a = 0.1" curve means the performance curves of the traditional protocols; "Theory" and "Simulation" curves denote the performance curves

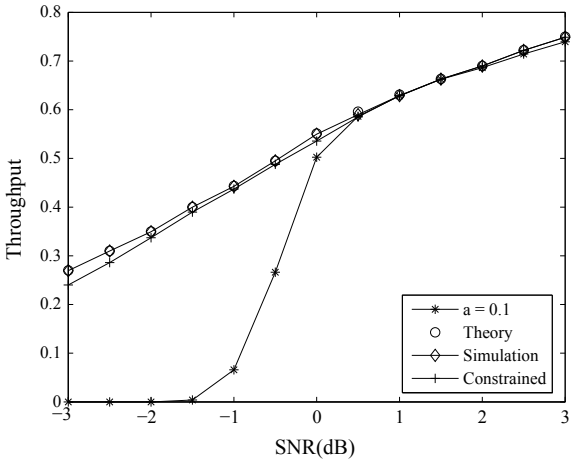


Fig. 3. The throughput performance of various RB-HARQ protocol.

of proposed protocol with maximum throughput which is calculated in theory and practice, respectively. "Constrained" curve indicates the performance of the scheme under constrain of the FER. From this figure, the actual simulation curve agrees with the theoretical curve.

As we analyzed, the RB-HARQ protocol with maximum throughout significantly outperforms all the traditional protocols at alternatively low SNR. When the schemes achieve the same throughput, the maximum throughput RB-HARB performs near 0.5 dB better than the RB-HARQ scheme under constrain.

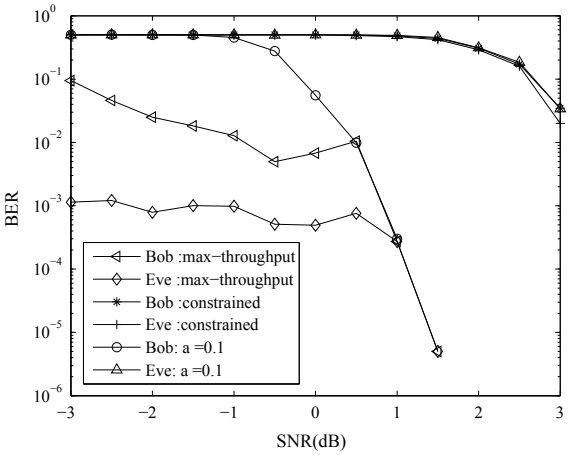


Fig. 4. BER performance of various RB-HARQ protocols

Figure 4 gives the BER performance of various RB-HARQs. At alternatively low SNR, the retransmission subsets change to increase the throughput, so the curve of proposed RB-HARQ waves irregularly. Unfortunately, the maximum throughput RB-HARQ faces a problem of reliability at relatively low SNR. But for the maximum throughput RB-HARQ with the constrain of FER, Bob's reliability performance is largely improved and keeps at relatively low level from the beginning. Compared with the maximum throughput RB-HARQ, the $SNR_{B,\min}$ of the scheme under constrain is improved by 3 dB.

5 Conclusion

In this paper, a RB-HARQ protocol combining with LDPC coding has been introduced into secure communication over AWGN channel. The protocol can adaptively compute the sizes of the retransmissions with the goal of optimizing throughput under the constrain of FER. The proposed RB-HARQ protocol has the potential of attaining large throughput and realizing reliable and secure transmissions.

References

1. Wyner, A.D.: The wire-tap channel. Bell Syst. Tech. J. **54**, 1355–1387 (1975)
2. Baldi, M., Bianchi, M., Chiaraluce, F.: Coding with scrambling, concatenation and HARQ for AWGN wire-tap channel: a security gap analysis. IEEE Trans. Inf. Forensics Secur. **7**, 883–894 (2012)
3. Tomasin, S., Laurenti, N.: Secure HARQ with multiple encoding over block fading channels: channel set characterization and outage analysis. IEEE Trans. Inf. Forensics Secur. **9**, 1708–1719 (2014)
4. Mheich, Z., Treust, M.L., Alberge, F., Duhamel, P.: Rate adaptation for incremental redundancy secure HARQ. IEEE Trans. Commun. **64**, 765–777 (2016)
5. Zou, Q., Zhang, B., Ma, Y., Guo, D.: Increasing physical layer security through reliability-based HARQ. In: 8th International Conference on Wireless Communications & Signal Processing, pp. 1–5. IEEE Press, NanJing (2016)
6. Szczecinski, L., Khosravirad, S.R., Duhamel, P., Rahman, M.: Rate allocation and adaptation for incremental redundancy truncated HARQ. IEEE Trans. Commun. **61**, 2580–2590 (2013)
7. Jabi, M., Pierre-Doray, E., Szczecinski, L., Benjillali, M.: How to boost the throughput of HARQ with off-the-shelf codes. IEEE Trans. Commun. **65**, 2319–2331 (2017)
8. Klinc, D., Ha, J., McLaughlin, S.W., Barros, J., Kwak, B.-J.: LDPC codes for the Gaussian wiretap channel. IEEE Trans. Inf. Forensics Secur. **6**, 532–540 (2011)
9. Visotsky, E., Sun, Y., Tripathi, V., et al.: Reliability-based incremental redundancy with convolutional codes. IEEE Trans. Commun. **6**, 987–997 (2005)
10. Moon, T.K.: Error Correction Coding: Mathematical Methods and Algorithms. Wiley, Hoboken (2005)
11. Wang, K., Ding, Z.: Diversity integration in hybrid-ARQ with chase combining under partial CSI. IEEE Trans. Commun. **64**, 2647–2659 (2016)

Research on Intelligent Service Aware Routing for Space-Ground Integrated Network

ChunFeng Wang(✉) and Naijin Liu

Qian Xuesen Laboratory of Space Technology, China Academy of Space Technology, Beijing, China
Jessen_wang@163.com, liunaijin@qxslab.cn

Abstract. Traditional routing protocols are used in Space-Ground Integrated Network will lead to more serious data loss, longer delay issues. This paper proposes an intelligent service aware routing method (ISAR). This method can use the local detection information to make the nodes quickly and accurately perceive the changes of the network environment and service status through the traffic flow perception and prediction, so as to optimize the network performance. ISAR uses the Q-learning algorithm to make the node have the ability of learning and reasoning, and improve the adaptability of the route to the service. The simulation results show that the method can effectively improve the utilization rate of the network link, realize the global network load balancing, improve the network throughput, and optimize the network performance.

Keywords: Space-ground integrated network · Intelligent service aware · Route optimization

1 Introduction

With the development of mobile Internet, the extensive application of information technology has had a profound impact on human society, the number of Internet users, also showed explosive growth. Recently, the traditional Internet model has been difficult to adapt to the growing needs of the network, but also cannot meet the user's new service needs for the network. In this process, the space–ground integrated network has gradually become the focus of next-generation network research. Space-ground Integrated Network is a large-capacity, multi-level heterogeneous network, can carry multiple information network [1]. As shown in Fig. 1, the Space-ground integrated network includes space satellite network and ground network. The space satellite network includes the GEO satellite backbone network and LEO satellite access network, covering a wide range of large-scale network; the ground network includes a variety of tandem station, control network, etc. and the service can be provided for pedestrian, car and ocean ship and other network services. Regional navigation enhancement, enhanced hotspot or emergency area communication, enhanced ground observation, ocean survey and navigation are also included. Space-ground integrated network has better coverage and the ground network has excellent transmission rate and scalability.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 462–472, 2019.
https://doi.org/10.1007/978-981-13-6264-4_56

In this dynamic and complex network, how to achieve the efficient use of resources to provide the QOS of service is still one of the hot areas of research. The space–ground integrated network is large, heterogeneous network coexistence and has more service types.

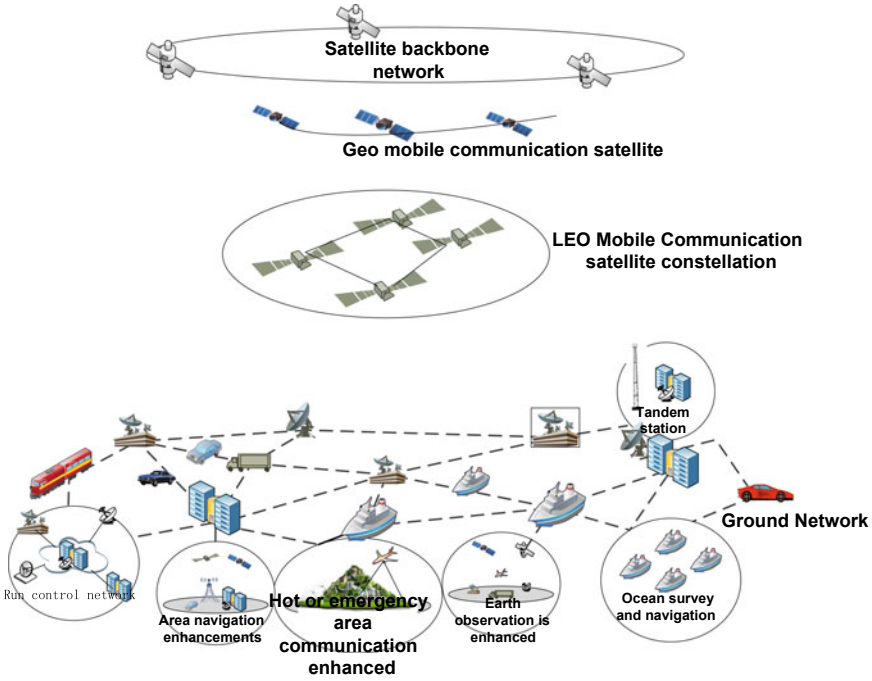


Fig. 1. Space-ground integrated network architecture

In this paper, we focus on that the heterogeneity and service complexity of space–ground integrated network leads to the problem that network performance is not high. By introducing SDN network concept [2–4] and architecture, and artificial intelligence method, the space–ground integrated network is optimized, which makes some nodes have perceptual and reasoning ability, By measuring or predicting the environmental parameters of the network, dynamic decision-making and network element reconstruction are realized, which can achieve the goal of network route adaptation and optimize communication performance. The remaining chapters of the paper are organized as follows.

The rest of this paper is organized as follows. Section 2 describes SDN-based integrated network architecture. Section 3 describes the intelligent service aware routing method system architecture and principle. Section 4 gives routing cost analysis. Section 5 is the simulation results. Section 6 is the conclusion.

2 SDN-Based Space–Ground Integrated Network Architecture

The routing nodes in space–ground integrated network are mainly composed of SDN core routers and edge routers. The space core router is loaded on the nodes of the GEO backbone network, while the space edge router is loaded on the LEO satellite access network node. As shown in Fig. 2, the edge router is to deploy the environment aware module and to identify the type of service from the current network environment and to collect network payload information. This information is uploaded to the SDN core router. SDN core router includes SDN controller module [5–7], SDN routing switch module, PCE routing computing service module. The core router implements data processing function, learning reasoning function, policy selection function, routing decision function and policy issuing function. The controller module through the edge router collection network topology information is transmitted to the core router and gives to PCE routing service module. SDN core router will analyze the collected information, reasoning, learning, routing decisions. The SDN controller will send route policy to the edge router; the edge router receives strategy service from the core router to distinguish different service and to implement intelligent management according to the routing strategy library content calculated by PCE calculation module of core routers as shown in Fig. 3.

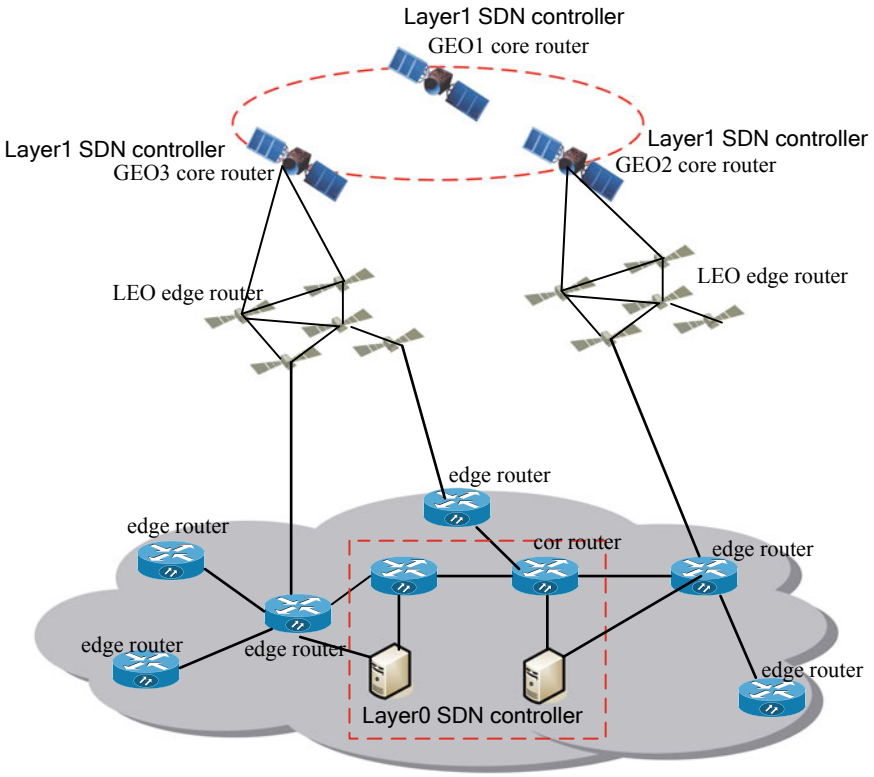


Fig. 2. SDN core router and edge router

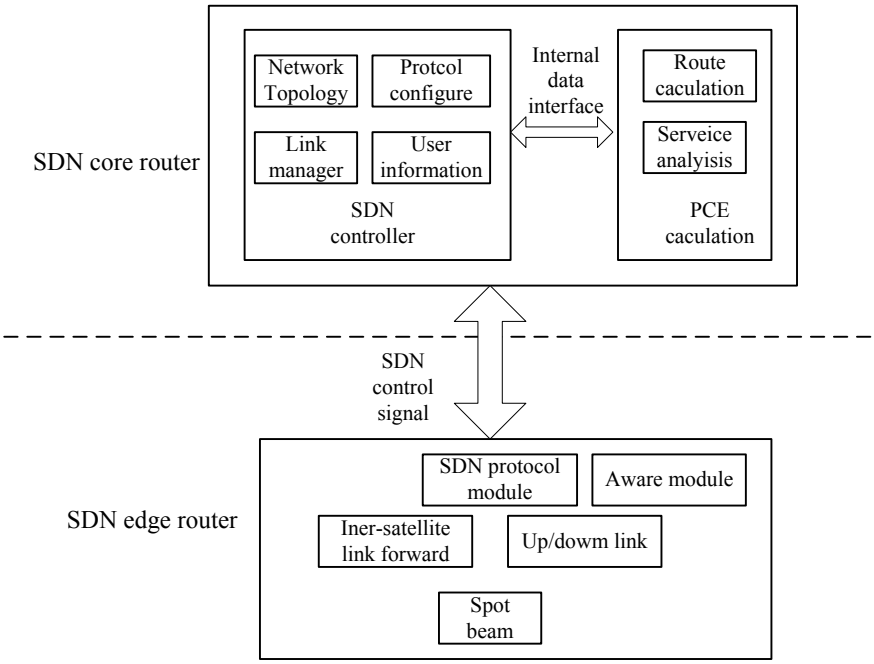


Fig. 3. SDN router module composition

3 Intelligent Service Aware Routing Network System

3.1 Intelligent Service Aware Routing System Architecture

As shown in Fig. 4, intelligent service aware routing system architecture can be divided into three layers: cognitive layer, learning layer and data layer. The data layer is responsible for encapsulating the data and data forwarding module. The learning layer includes routing strategy performance evaluation module, policy selection algorithm and policy library update module, route prediction module. The cognitive layer includes QOS aware module, route discovery module, routing decision module, route configuration module. The node can obtain the network environment measurement information through the collection of user service request and feedback and through network monitoring and other means, through the routing strategy selection and other means to achieve route reconstruction, so as to ensure the high performance of the network.

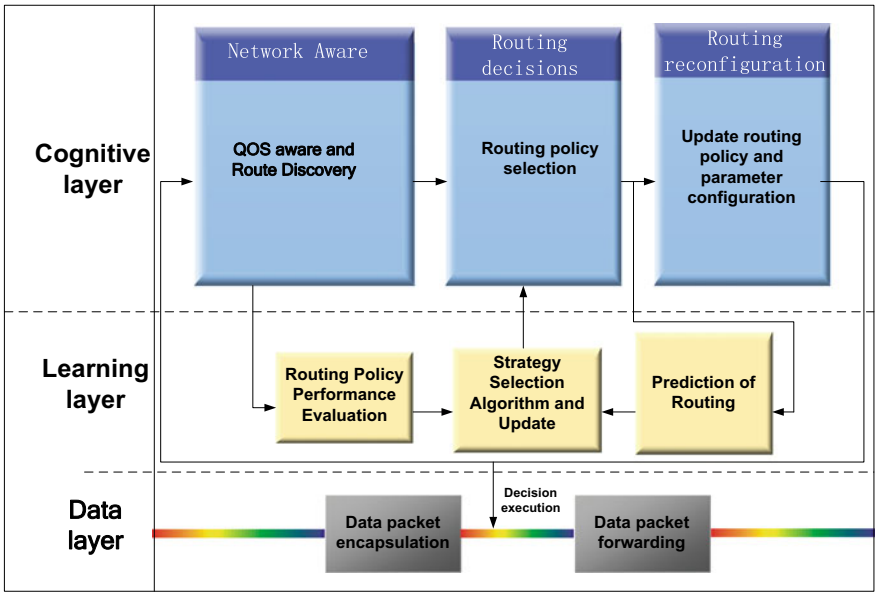


Fig. 4. Intelligent service aware routing system architecture

(A)
The Cognitive layer

The Cognitive layer includes network aware, routing decisions, and routing reconfiguration.

The network-aware module has two main functions, namely, QOS service aware function and network routing discovery function. The QOS service aware function maps the demand for the service flow to the QOS requirements in the network by collecting user service requests and feedback information. The network routing discovery function obtains the route information through the route protocol such as LAOR protocol [8] and MOR protocol [9], REQ is Route Requirement message.

Routing decision module is mainly responsible for the construction of the route, update and so on. It is based on network aware information select the routing strategy, such as multi-path routing and cross-layer routing.

The routing reconfiguration module is responsible for the update configuration of the route. If you choose a cross-layer routing policy, you need to update the configuration physical layer, link layer, and network layer.

(B) **The Learning layer**

The main functions of the learning layer are the performance evaluation of the routing strategy; the updating of the strategy selection algorithm and the strategy library, the routing strategy prediction.

(C) **The Data layer**

The data layer is on the bottom of intelligent service aware routing. The data layer is the physical node of the network, in the network reconstruction and routing, the collection of network topology, link status, load conditions, and the current environmental conditions and other information will be provided to the Learning layer, and through the cognitive layer to complete the routing strategy sent to the data layer to perform. According to service needs and network environment information and according to the routing strategy it is to implement network resource allocation and data forwarding.

3.2 Service Class

Space–ground integrated network includes elastic and non-elastic data and the elastic data is such as FTP and other traditional data services and the non-elastic data is such as video, Live, VoIP and other multimedia services. The two types of data have different requirements on the data transmission capacity and quality. so the network is considered to support three different service classes: Class A, typical real-time services such as Voice over IP (VoIP) and interactive video applications. This traffic class is sensitive to packet delay, which needs to be minimized; Class B, best-effort traffic, which has no specific QOS constraints; Class C, bandwidth-sensitive services such as video on demand and distribution of large files and This traffic class requires throughput to be maximized. It is necessary to consider how to allocate the current network resources effectively to realize the efficient transmission of the network and make full use of the existing limited resources to meet the users' demand for data transmission services. In the network and between users, it is to achieve an effective balance of network service capabilities, give full consideration to the use of network resources and the fair use of network resources.

3.3 Service Traffic Characteristic Matrix

Network services can be described with six parameters such as Service Flow Category, denoted by C, the source address, denoted by S, the destination address, denoted by D, the bandwidth requirement, denoted by B, the node one hop delay, denoted by T, the node packet loss rate, denoted by P_{Loss}. The six-tuple array can be expressed as the network traffic characteristic matrix M (C, S, D, B, T, P_{Loss}). If the service flow is classified into K class by service sense, C_k is the type of service flow and B_k is the bandwidth requirement of the k class service flow, respectively. In this way, the current network traffic flow type, distribution, bandwidth requirements, delay packet loss rate characteristics can be expressed using the network traffic feature matrix *M*.

3.4 Learning Process

At present, in the study of space–ground integrated network research, most of the routing protocols will not change the path parameters after the path is established. This makes the routing protocol difficult to adjust and optimize according to the network

environment. In the ISAR method, Q-learning algorithm [10, 11] is used to improve network performance. Q-learning is a kind of reinforcement learning, so that the system can continue to use feedback to select the optimal action to achieve its goal. In the ISAR method, the following objective function is defined for Q-learning [12]:

$$G = (1 - P_{loss})T + P_{loss}(t + G) \tag{1}$$

where P_{LOSS} is the average packet loss rate of the system, T is the hop-by-hop average delay; t is the time to wait for retransmission after packet is loss. In Q-learning, the reward is R and R = 1/G. Each node is to learn according to their own link state parameter changing, each change immediately get a reward R, if R is better than the previous reward value, indicating that the network to change in a good way. R(S, p) is the reward function. We have the following reward function matrix:

$$\begin{pmatrix} R(S_1, p_1) & \cdots & \\ & \ddots & \vdots \\ & \cdots & R(S_n, p_n) \end{pmatrix} \tag{2}$$

where $S_1 \ldots S_n$ is network topology node, $p_1 \ldots p_n$ is the link state parameter in the network topology. By calculating from the formula (1), the transfer rule matrix can denoted as the following formula (3):

$$Q(s, p) = R(s, p) + \gamma \cdot \max\left\{Q\left(s', p'\right)\right\}. \tag{3}$$

In the formula (3), γ is the learning factor in the Q-learning algorithm. Use the following algorithm 1 to train the network:

Algorithm 1:
Step 1: Given the parameter γ and the reward matrix R.
Step 2: Initialize Q: = 0.
Step 3: For each episode:

Randomly select an initial state S.
If the target state is not reached, perform the following steps

(1) Select A action in all possible action of the current state s.
(2) Using the selected action A, to get the next state s′.
(3) Calculate Q (s, p) according to formula (3).
(4) Set s: = s′.

3.5 Strategy Selection

Because that the satellite network node have complex network terminals and link characteristics, this will lead to packet loss rate and link attenuation rate surge, and also there will be a bad space environment affecting satellite communications, such as solar wind, space particle Interference and so on. Therefore, in the choice of routing path, the

global optimal path should be selected. Through intelligent service aware routing for space–ground integrated network, which implement by using the core router routing decision through the appropriate path algorithm, the edge router collecting data, and receiving the issued routing table and forwarding the data, can effectively improve the network flexibility and survivability.

Space–ground integrated network carries a variety service of different bandwidth requirements. Because of the bad environment in the routing node, some bandwidth may not be used, in order to effectively use the network source; the optimal network bandwidth allocation strategy selection is needed. The policy selection is based on the end to end QOS target, the network survival conditions, the type of service in the network, the network traffic state, and so on. In order to effectively use the limited network source, making the traffic flow can be distributed in different links; it is needed as far as possible to choose the link capacity and traffic bandwidth demand ratio and the traffic flow and link bandwidth is used to control the available bandwidth allocation in the network.

4 Cost Analysis

In order to further understand the performance of the ISAR routing, the cost of the method is analyzed. Table 1 gives some of the symbols and definitions that will be used below.

Table 1. Symbol defines

N	Node number
L	Average path length
$\boldsymbol{\lambda_R}$	Average number of REQ packets per node per second
$\boldsymbol{B_{REQ}}$	REQ packet size
$\boldsymbol{\lambda_p}$	The number of packets per node per second
$\boldsymbol{B_p}$	Packet size
$\boldsymbol{P_{LOSS}}$	Average packet loss rate per node
$\boldsymbol{H_k}$	The number of hops passed before the packet is discarded

4.1 Control Overhead

N is the number of nodes in the network, the average number of packets per second network REQ packet is $\lambda_R N_\circ$ REQ packet flooding up to $N - 1$ times. The overhead of REQ packets in the whole network is $O(\lambda_R N^2 B_{REQ})$.

4.2 Packet Loss Overhead

When the source node establishes a path with the destination and sends data packets, if the packet is discarded in the transmission of H_k hop before to the destination node, the routing overhead transmission of the data packet is the packet loss caused by H_k. The

probability that each node discard packet is P_{LOSS}, and the probability that packets will be dropped at K hop is $P\{X = k\} = (1 - P_{LOSS})^{k-1}(P_{LOSS})$, k = 0, 1, 2, 3, k = 0, 1, 2, 3… The average number of hops for packet loss is $X = P_{LOSS} \sum_{k-1}^{L-1} (1 - P_{LOSS})^{k-1}$. The average per packet overhead caused by packet loss in a transmission procedure is EXB_p, the average number of retransmission is $1/(1 - P_{LOSS})^L$. The whole network creates $\lambda_t N$ packet every second, the overhead because of the packet loss is $O\left(\lambda_p N P_{LOSS} \sum_{k-1}^{L-1} (1 - P_{LOSS})^{k-1} \Big/ \left((1 - P_{LOSS})^L B_p\right)\right)$.

5 System Simulation and Results Analysis

5.1 Simulation Principle and Environment

In order to evaluate the performance of ISAR in the two aspects of data packet loss and resource utilization for space–ground integrated network, the simulation software system is developed based on C++. The software system includes network event module, network topology module, network resource module, service generator module, simulation control and statistics module.

The simulation example includes six nodes. There are five satellite node and one ground communication node. The node1, node2 and node3 are GEO satellite. The node4 and node5 are LEO satellite. The node6 is the ground PC communication node.

5.2 Simulation Results and Analysis

The simulation results are shown in Figs. 5 and 6. The simulation results show that the ISAR method will improve network performance and reduce network block rate and improve the utilization of network resources.

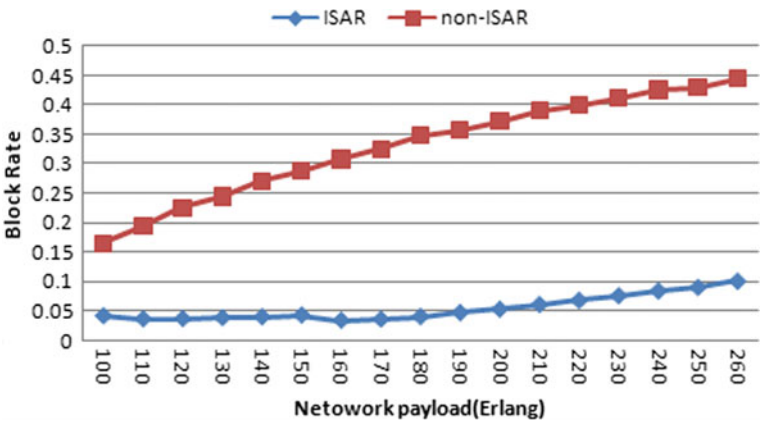


Fig. 5. Block rate for ISAR and non-ISAR

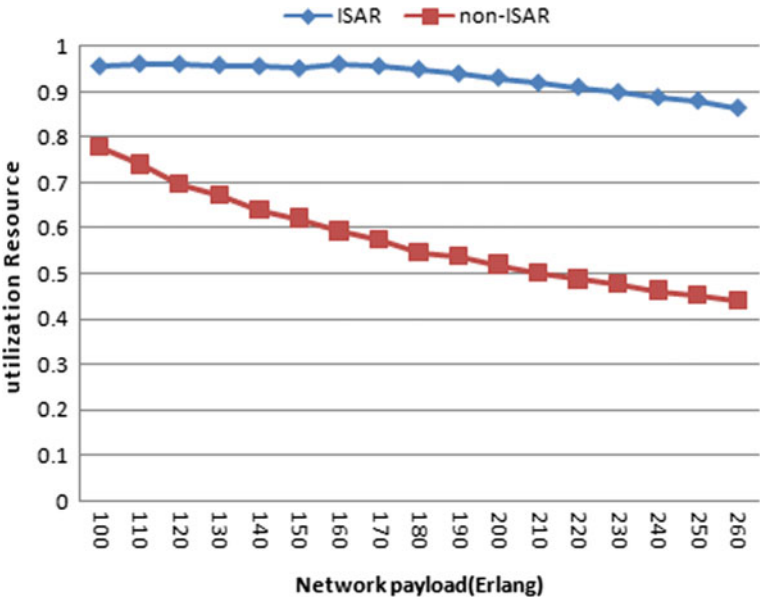


Fig. 6. Resource utilization for ISAR and non-ISAR

6 Conclusion

This paper proposes an intelligent service aware routing method based on cognitive learning for space-ground integrated network. The space-ground integrated network architecture based on SDN is shown. The system architecture and principle of intelligent service aware routing method is described in detail. Intelligent service aware routing method includes three layers, respectively is the cognitive layer, learning layer, data layer, this paper expounds the learning process of Q-Learning algorithm for network service analysis, the overhead of the routing method. Finally, the system simulation is done, the simulation results show that the intelligent service aware routing method reduces the network congestion rate, improve the network source the utilization rate, and can effectively improve the network performance.

References

1. Jianing, G., Changde, L., Yun, S.H.I., Bin, W.U.: Satellite-ground integrated communication, network based on distributed data center. Chin. Space Sci. Technol. **35**(5), 77–84 (2015)
2. OpenFlow Switch Specification Version 1.4.0. https://www.opennetworking.org
3. Shin, S., Porras, P., Y egneswaran, V., et al.: FRESCO: modular composable security services for software defined networks. In: Proceedings of the ISOC Network and Distributed System Security Symposium, San Diego, CA, February 2013

4. Porrasy, P., Shinz S., Yegneswaran V., et al.: A security enforcement kernel for OpenFlow networks. In: Proceedings of the ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking (HotSDN), Helsinki, Finland, August 2012
5. Hassas Yeganeh, S., Ganjali, Y.: Kandoo: a framework for efficient and scalable offloading of control applications [C]. In: Proceedings of the First Workshop on Hot Topics In Software Defined Networks, pp. 19–24. ACM (2012)
6. Sooel, S., Seungwon, S., Vinod, Y., et al.: Model checking invariant security properties in openflow. In: Proceedings of IEEE International Conference on Communication (ICC). IEEE Press, Budapest (2013)
7. Reitblatt, M., Foster, N., Rexford, J., et al.: Abstractions for network update. In: Proceeding of the SIGCOMM 2012, pp. 323–334. ACM Press, Helsinki (2012)
8. Papapetrou, E., Karapantazis, S., Pavlidou, F.N.: Distributed on-demand routing for LEO satellite systems. Comp. Netw. **51**(15), 4356–4376 (2007)
9. Karapantazis, S., Papapetrou, E., Pavlidou, F.-N.: Multiservice on-demand routing in LEO satellite networks. IEEE Trans. Wirele. Commun. **8**(1), 107–112 (2009)
10. Muñoz, P., Barco, R., de la Bandera, I.: Optimization of load balancing using fuzzy Q-learning for next generation wireless networks. Expert Syst. Appl. **40**(4), 984–994 (2013)
11. Galindo-Serrano, A., Giupponi, L.: Distributed Q-learning for aggregated interference control in cognitive radio networks. IEEE Trans. Vehicular Technol. **59**(4), 1823–1834 (2010)
12. Gelenbe, E., Gellman, M., Lent, R., et al.: Autonomous smart routing for network QoS. In: Proceedings of the First International Conference on Autonomic Computing, pp. 232–239. IEEE Computer Society, New York (2004)

Efficient Resource Allocation Algorithm to Minimize Interference Among Co-channel Users in Ultra-Dense Networks

Yutong Xiao(✉), Xuanli Wu, Chuiyang Meng, and Wei Wu

Communication Research Center, Harbin Institute of Technology, Harbin, China
{hitadao,mcymcymcymcy}@163.com, {xlwu2002,kevinking}@hit.edu.cn

Abstract. Ultradense network (UDN) is an effective solution to providing high capacity and massive connections in hotspots. However, a large number of randomly deployed small base stations (SBSs) cause severe interference. In order to solve the strong co-channel interference among user equipment (UE), we proposed an efficient resource allocation algorithm based on simulated annealing (SA). First, we established a system-level architecture of UDN, which is more realistic and suitable for dense SBSs scenario. In addition, considering the properties of UDN, we formulated the resource allocation problem into a 0–1 problem. Moreover, we transform this problem into a traveling salesman problem (TSP) and propose a resource allocation algorithm to solve it. Unlike the previous works, we study different ratios between UEs and SBSs. Simulation results show that the proposed algorithm can efficiently avoid co-channel interference and improve system throughput with lower signaling overhead. Moreover, SBSs multiple association is excellent in improving system throughput.

Keywords: Ultradense network · Resource allocation · Simulated annealing · Traveling salesman problem

1 Introduction

The fifth generation mobile communication systems are expected to provide tens of Gbps data rates and support 10,000 devices [1]. One effective solution is to increase the density of small base stations (SBSs) in ultradense network (UDN) system. High density deployment of SBSs can shorten the distance between SBS and user equipment (UE), improve the communication quality and enhance regional spectrum efficiency. However, mutual co-channel interference will be stronger than ever before and diminish the system performance gains, as the numbers of UEs and SBSs are large.

Ideal resource allocation can effectively tackle the issue of co-channel interference. The authors in [2] compared the performances of proportional fairness (PF) and round robin (RR) schedulers in dense networks. They found that RR has a similar performance as PF. However, the system performance has much room for improvement. The existing resource allocation approaches cannot be directly applied in UDN, as the resource allocation problem is always non-convex and NP-hard, which is difficult to

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 473–480, 2019.
https://doi.org/10.1007/978-981-13-6264-4_57

find an optimal solution. Furthermore, high computational complexity and overhead make these approaches difficult to use in real scenario.

A divide and conquer strategy has been proposed in current research [3, 4], which divide SBSs into several clusters to simplify large-scale resource allocation problem. However, due to the randomness of massive SBS locations, it is difficult to find a criterion of clustering. Especially when SBSs are dense enough, all SBSs may belong to the same cluster and in different areas, there will be different criterions.

Hierarchical strategy has also been conceived as an effective method to decompose the original resource allocation problem. In [5], the authors divide the resource allocation problem into four stages with partially distributed management. The work in [6] transforms the resource allocation problem into a two-stage Stackelberg game. Nevertheless, it is difficult to decompose the original problem into several disjoint subproblems.

Several studies adopt game theory to obtain optimal solution [6, 7], while others convert the non-convex resource allocation problem into the solvable forms [8]. However, these methods need multiple iterations, which require much time in large-scale networks. Resource allocation based on graph coloring is introduced in [9, 10]. However, when the network is dense enough, the conflict graph is extremely complex. Besides, the graph is time-varying with the movement of UE. Moreover, these studies only considered that UEs number is less than or equivalent to SBSs.

In this work, we propose an efficient resource allocation algorithm to mitigating co-channel interference of UEs. The resource allocation problem is a 0–1 problem, which is NP-hard in UDN. Different from the aforementioned existing works, we model the problem into a modified traveling salesman problem (TSP) and we propose a resource allocation algorithm based on simulated annealing (SA) to solve it. We also analyze the performance of the proposed algorithm with different ratios between UEs and SBSs, which is important but is ignored in many studies.

The remainder of this paper is organized as follows: Sect. 2 describes the system model and formulates the resource allocation problem of UDN. In Sect. 3, we transform the problem into a traveling salesman problem and introduce the SA algorithm to solve it. Section 4 presents simulation results and discussion. Finally, conclusions are drawn in Sect. 5.

2 System Model and Problem Formulation

In this paper, we focus on the uplink scenario, where UE's transmit power is small and has little difference in UDN [11]. Macro base station (MBS) manages the whole system and SBSs provide high-quality service for small cell UEs (SUE). MBS serves macro user equipment (MUE), which cannot be served by SBSs. MBS and SBSs work on different spectrum, which means the MUE and SUE will not interfere with each other. It is assumed that the coverage areas of MBS and SBS are round, with radius r_m and r_s, respectively. Depending on UE's traffic and the mean received signal strength measurement, SUE decides which SBS to connect with. UEs can connect up to M SBSs. When $M = 1$, the UE is in single connection association, and $M > 1$ is multiple association.

We adopt a modified Poisson point process (PPP) model to imitate the deployment of UEs and SBSs with the density of λ_U and λ_B, respectively. We consider different ratios $r = \lambda_U/\lambda_B$ between UEs and SBSs. Besides, we consider more distance restrictions between UEs-SBSs, UEs-MBS, and SBSs-MBS in the site model, while others only considered single restriction among SBSs or UEs. With these restrictions, the site model is more approximate to the realistic scenario.

UDN evolves from heterogeneous network (HetNet), and the 3rd Generation Partnership Project (3GPP) recommends a LOS/NLOS path loss model which is suitable for dense heterogeneous networks and hotspots in outdoor communication environments [12, 13]. By comparing UDN path loss model with LTE in Fig. 1, we concluded that the probability of using LOS propagation in UDN is bigger than LTE system. And UDN path loss is always bigger than LTE.

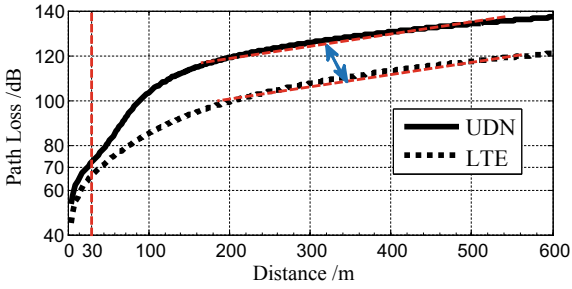


Fig. 1. Path loss of UDN and LTE system

We define $R = \{rb_i | i = 1, 2, 3, \cdots, m\}$ as the set of available resource blocks; the set of SUE is $S = \{sue_j | j = 1, 2, \cdots, n\}$. We use $X_{n\times n}$ to denote the RB's allocation situation for SUEs. $x_{i,j} \in \{0, 1\}$ is the i-th row and j-th column of X. If rb_i is allocated to sue_j, then $x_{i,j} = 1$. For sue_j, its association SBS set is $B_j = \{b_{jk} | k = 0, 1, 2, 3\}$. The SUEs which allocated the same RB i are $I^i = \{u_a^i | u_a^i \in \mathrm{S}\}$. Let $d_{j,p}$ represent the distance between sue_j to SBS bs_p^j.

The resource allocation problem in UDN can be modeled as a 0–1 problem in integer programming:

$$\begin{aligned} \max F = & \sum_{i=1}^{m} \sum_{j=1}^{n} x_{i,j} \cdot X_{i\cdot} \cdot d_{\cdot f} \\ s.t. \ & \sum_{i=1}^{m} x_{i,j} = 1 \\ & x_{i,j} = 0, 1 \end{aligned} \tag{1}$$

where $X_{i\cdot}$ is the i-th row of X and it records the UEs which get rb_i, $d_{\cdot f}$ denotes the f-th column of D, which records distances among all UEs to ue_f. Constraints indicate that in each slot we allocate resources and each UE will get one RB. At present, Eq. (1) has not been solved in an effective way.

In this section, we adopt a modified PPP model to simulate the UEs and SBSs deployment process, which approximates the realistic scenario. Besides, we study the

path loss model that satisfies the SBSs densification scenario. By analyzing the interference caused by UEs which get the same frequency resource, the problem is formulated as a 0–1 problem. It is difficult to find the optimal solution in polynomial time.

3 Resource Allocation and Traveling Salesman Problem

In UDN, there are many UEs and SBSs which make the complexity of the resource allocation problem is greater than ever before. If neighboring UEs get the same spectrum resource, they will cause strong interference to each other. Many works adopt game theory or convex optimization theory to solve the problem. However, the signaling overhead and computational complexity will restrict their effects in practical applications. In order to avoid these problems, we transform the RA problem into a modified TSP, which has many similarities with the RA problem we formulated. And heuristic algorithm can effectively solve TSP with lower signaling cost and more effective complexity control.

A brief description of TSP is that a sale man is going to some cities to sell products and the distances between any two cities are known. The sale man needs to visit all cities and get back to the starting city at last. The aim is to find the permutations of all cities, which will minimize the route length. Denotes the solution space of TSP is

$$S_{TSP} = \{(c_1, c_2, \cdots, c_n)\} \tag{2}$$

where n is the number of cities; $c_i = city_l$ represents the i-th arrived city is $city_l$; $(c_1, c_2, \cdots, c_n)$ is the permutation of $\{1, 2, \cdots, n\}$. The objective function is the route length of all cities:

$$C(c_1, c_2, \cdots, c_n) = \sum\nolimits_{i=1}^{n-1} d(c_i, c_{i+1}) + d(c_1, c_n) \tag{3}$$

where $d(c_i, c_{i+1})$ is the distance between the i-th arrived city and the $(i+1)$-th arrived city. As the objective function (1) is non-convex, there are many local optimal solutions.

At first, we imitate TSP to build a "Candidate City" set. In our modified model, the number of RB is limited, each RB will be allocated $Q = N_{sue}/N_{rb}$ times, where N_{sue} and N_{rb} are the numbers of SUE and RB, respectively. Then the "Candidate City" set can be denoted as

$$RB' = \{rb_1^1, rb_1^2, \cdots, rb_1^Q, rb_2^1, rb_2^2, \cdots, rb_2^Q, \cdots, rb_{N_{rb}}^1, rb_{N_{rb}}^2, \cdots, rb_{N_{rb}}^Q\} \tag{4}$$

The solution is a non-repeating permutation set $S_{RA} = \{(c_1', c_2', \cdots, c_{N_{sue}}')\}$ from the "Candidate City". And $c_i' = j$ represents the i-th SUE gets the j-th element in the set RB'.

SA algorithm has more advantages in actual engineering realization and complexity control in TSP. We can adjust the parameters of SA to trade off performance and complexity. Besides, the SA algorithm is suitable for finding the optimal solution of a problem

with large feasible solutions [14]. The initial resource allocation solution is the result of RR scheduling, which guarantees the performance of SA is no less than RR. Note that SA can calculate the results in each time slot, so it also applies to mobile users.

4 Simulation Results and Discussion

In UDN system, the dense deployment of SBSs can provide a higher preference of the system throughput and boost transmission rate. However, as the numbers of SBSs and UEs increases, serious interference will degrade the system performance and make the management and maintenance of the system more challenging. We transform the resource allocation problem into a modified TSP. Then, we propose a new and effective resource allocation algorithm based on SA to solve it. MBS only needs the locations information and the connection status of UE-SBS pairs, the algorithm can avoid high signaling cost.

In order to evaluate the performance of the proposed algorithm, we compare SA with RR scheduling in single and multiple association scenarios. Figure 2 shows the results with the ratio $r = 1$ between UEs-SBSs. When the SBS density is increasing, the system throughput is also increasing. The system throughput of SA is always bigger than RR scheduling. Multiple association performance grows faster than single association, and SA with multiple association is more likely to get a linear gain of throughput in UDN.

Continuing to increase the number of base stations cannot improve SBSs' utilization and the proportion of SUEs. When the number of SBS is 400, the system throughput is no longer growing at the final for single association with RR. Its density is given by

$$\lambda_B = \frac{400\text{cells}}{\pi(300\text{m})^2} \approx 1415\,\text{cells/Km}^2 \tag{5}$$

The number of UEs is dynamical change in UDN system. Figure 3 shows the situation where the ratio between UEs and SBSs is $r = 1/3$, the curves trend is similar to Fig. 2. In this situation, the proportion of SBSs which only serve one UE is increasing. The multiple association is excellent in improving system throughput.

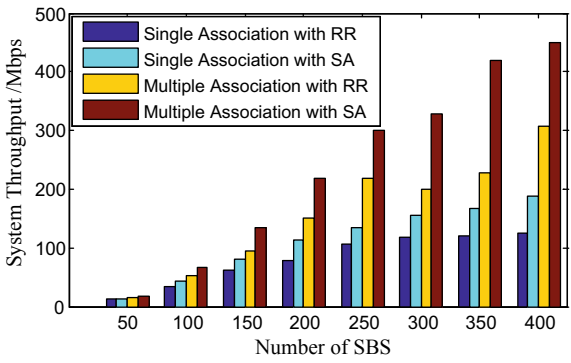


Fig. 2. The number of SBSs is equivalent to the number of UEs

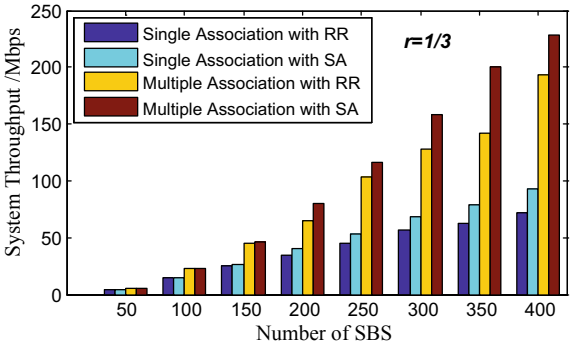


Fig. 3. The number of SBSs is less than the number of UEs

Considering the fact that we set $M = 3$, multiple association throughput is expected 3 times of single association. However, the goal is not achieved until SBSs number reaches 350 with SA, while it is 400 with RR. The reason is that when SBSs density is small, some UEs cannot be served by M SBSs or even be served by MBS.

When UEs number is bigger than SBSs, the interference among UEs is very serious. Moreover, the performance gain obtained by UEs from multiple association is also limited. Figure 4 presents the system throughput with $r = 3$. When SBSs number is 100 at Fig. 4 and SBSs number is 300 in Fig. 2, the system performance of $r = 3$ is worse than $r = 1$. And the advantage is not obvious for multiple association. It is obvious that the system performance does not have an advantage when r is big. It is better to add more SBSs in the system.

As for average throughput of each UE, in Fig. 5, it has been declining in single association with RR. However, the average throughput will first increase and then decline in other situations. The turning point is 150 of SBSs number. Moreover, the curves in Figs. 4 and 5 fluctuate. This is because of the random deployment of SBSs and UEs. The more UEs in the system, the stronger fluctuations of the system performance are.

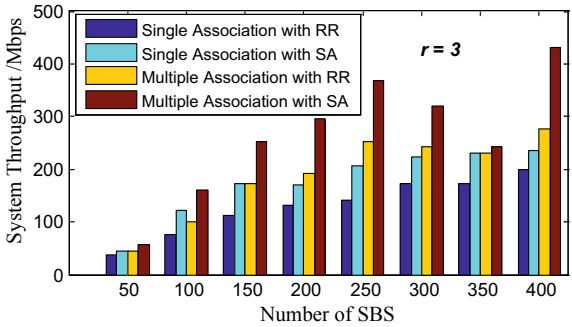


Fig. 4. The number of SBSs is bigger than the number of UEs

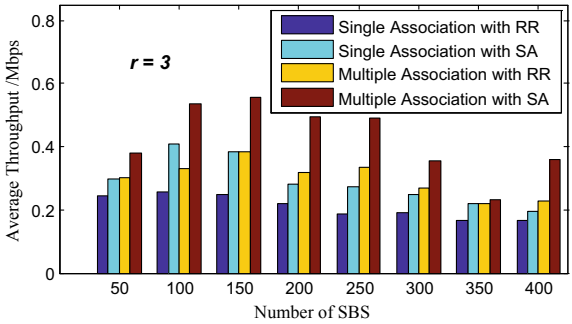


Fig. 5. Average throughput of UEs

The algorithm we have proposed can efficiently improve system throughput by maximizing distances among UEs which are allocated the same RB. When we combine SA with multiple association, the performance gain is very significant, especially when r is smaller than 1. When the number of SBSs is bigger than UEs, it will introduce stronger interference and may make UEs average throughput decline. Therefore, considering the value of M, it is better to keep $r = 1/M$ by adding or turning off SBSs in the system dynamically, while keeping the SBSs density below 1415 cells/Km^2.

5 Conclusion

In this paper, aiming at improving system capacity and mitigating co-channel interference in UDN system, we proposed a resource allocation algorithm based on SA to maximize the distances among UEs which are allocated the same RB. We transformed the resource allocation problem into a modified traveling salesman problem. Considering UDN's constraints and engineering realization requirements, we adopt the SA algorithm to solve the problem. According to the number of UEs, SA can adjust its parameters dynamically and flexibly to trade off the performance with computational complexity. We also studied the effect of different ratios between UEs and SBSs on system performance. The simulation shows that the proposed algorithm can effectively mitigate the interference and increase the system throughput in both single and multiple association scenarios, up to 54% and 84% when the ratio of UEs to SBSs is 1, respectively. Besides, multiple association with SA algorithm can greatly make the system performance increase when the number of SBSs is bigger than the number of UEs and keeping SBS's density below 1415 cells/Km^2.

Acknowledgements. This research is supported by Natural Science Foundation of China (Grant No. 61501136).

References

1. Andrews J.G., et al.: What Will 5G Be? IEEE J. Sel. Areas Commun. **32**(6), 1065–1082 (2014)
2. Jafari, A.H., Lopez-Perez, D., Ding, M., Zhang, J.: Study on scheduling techniques for ultra dense small cell networks. In: 2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall), Boston, MA, pp. 1–6 (2015)
3. Zhou L., et al.: A dynamic graph-based scheduling and interference coordination approach in heterogeneous cellular networks. IEEE Trans. Vehicular Technol. **65**(5), 3735–3748 (2016)
4. Liang, L., Wang, W., Jia, Y., Fu, S.: A cluster-based energy-efficient resource management scheme for ultra-dense networks. IEEE Access **4**, 6823–6832 (2016)
5. Qiu, J., et al.: Hierarchical resource allocation framework for hyper-dense small cell networks. IEEE Access **4**, 8657–8669 (2016)
6. Liu, Y., Wang, Y., Zhang, Y., Sun, R., Jiang, L.: Game-theoretic hierarchical resource allocation in ultra-dense networks. In: 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, pp. 1–6 (2016)
7. Qiu, J., Wu, Q., Xu, Y., et al.: Demand-aware resource allocation for ultra-dense small cell networks: an interference-separation clustering-based solution[J]. Trans. Emerg. Telecommun. Technol. **27**(8), 1071–1086 (2016)
8. Luo, Y., Hua, C.: Resource allocation and user-centric clustering in ultra-dense networks with wireless backhaul. In: 2016 8th International Conference on Wireless Communications & Signal Processing (WCSP), Yangzhou, pp. 1–5 (2016)
9. Zhao, L., Wang, H., Zhong, X.: Interference graph based channel assignment algorithm for D2D cellular networks. IEEE Access **6**, 3270–3279 (2018)
10. Yang, T., Zhang, R., Cheng, X., Yang, L.: Graph coloring based resource sharing (GCRS) scheme for D2D communications underlaying full-duplex cellular networks. IEEE Trans. Vehicular Technol. **66**(8), 7506–7517 (2017)
11. Onireti, O., Imran, A., Imran, M.A., Tafazolli, R.: Energy efficient inter-frequency small cell discovery in heterogeneous networks. IEEE Trans. Vehicular Technol. **65**(9), 7122–7135 (2016)
12. 3GPP TR 36.814 V1.2.1. (2009). Further advancements for EUTRA: Physical layer aspects, Tech. Specification Group Radio Access Network Rel. 9, June 2009
13. Galiotto, C., Pratas, N.K., Marchetti, N., Doyle, L.: A stochastic geometry framework for LOS/NLOS propagation in dense small cell networks. In: 2015 IEEE International Conference on Communications (ICC), London, pp. 2851–2856 (2015)
14. Ingber, L.: Simulated annealing: practice versus theory. Math. Comput. Modell. Int. J. **18** (11), 29–57 (1993)

PLDRA: Probability-Based Loss Differentiation Rate Adaptive Algorithm for IEEE 802.11p

Jingtao Du, Shubin Wang(✉), Minghao Wang, and Caiqing Wang

College of Electronic Information Engineering, Inner Mongolia University, Hohhot 010021, China
wangshubin@imu.edu.cn

Abstract. The traditional rate adaptive algorithm based on channel detection fails to work well under the vehicular network of which channel quality changes quickly because of the fixed threshold of rate switching. This paper proposes a rate adaptive algorithm with loss differentiation for vehicle network based on the probabilities of which kinds of frame lost (PLDRA). This paper proposes a suitable threshold of SNR-MCS switching for IEEE 802.11p under VANET. And by comparing the size of frame error rate and collision rate as the basis of packet loss, the next transmission rate is selected by using both the threshold and situation of packet loss. Simulation results show that the PLDRA algorithm can achieve higher accuracy for channel quality prediction and higher system throughput under VANET, compared with ARF algorithm and FARA algorithm.

Keywords: VANET · IEEE 802.11p · Rate adaptive · SNR · Loss differentiation

1 Introduction

Vehicular ad hoc network (VANET) is a special mobile ad hoc network which can achieve vehicle-to-vehicle communications and vehicle to roadside unit communications. Due to the fast changing and complex channel environment, the wireless transmission efficiency of the vehicle network system is not good enough [2].

The frequent packet loss in the IEEE 802.11 WLAN is due to the high bit error rate of the wireless link. However, for each packet, there are mainly two types of packet losses. One of the types is due to collisions, which is called collision packet loss; the other type is the decoding error which is caused by weak signals; and it is called weak signal packet loss. The reason for packet loss reflects the current wireless channel status. However, the standard of IEEE 802.11 does not consider the reason for distinguishing between packet loss or channel error, and adopts a conservative solution to packet loss. The program works well in the static channel parameters, but without considering the node's mobility at the design stage suffers from mobility [1]. Weak signal packet loss with the mobility is caused by the over-optimization of rate.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 481–489, 2019.
https://doi.org/10.1007/978-981-13-6264-4_58

Although IEEE 802.11 provides multiple transmission rates at the physical layer, the mac layer does not specify how to select an appropriate transmission rate. The earliest and the most current hardware vendor still adopts the adaptive rate algorithm ARF (auto rate fallback) algorithm [5]. If there are 10 successfully transmissions in succession, it indicates that the channel quality is good and attempts to use a higher transmission rate. If there are two consecutive transmissions fail, it will switch to a low rate. The ARF algorithm is easy to implement without the requirement to modify the mac frame format. There are other kinds of rate adaptation based on channel detection. Reference [8] presents FARA (frequency-aware rate adaptation), which presents a scheme to robustly compute per-frequency SNRs using normal data transmissions. The threshold of SNR-MCS switching is generated by the WiGLAN radio platform running all possible bit rates for the whole SNR range to get the minimum SNR required to maintain the packet loss rate below 1%. The SNR-MCS table determines the performance of the algorithm, but it is designed for IEEE 802.11a.

The above two methods do not distinguish packet loss, the existing research on the cause identification of WLAN error codes packet loss is mainly based on the sender identification. The literature [7] proposes a method for identifying the error causes of exchanging RTS (request to send)/CTS (clear to send) control messages between sender and receiver is proposed. The literature [6] proposes a method combining RTS/CTS with data segmentation to distinguish error codes and congestion packets. All of these methods need to exchange RTS/CTS control packets many times, and the communication cost is high. In literature [9], a COLLIE discriminant method is proposed, which mainly determines the cause of packet loss by analyzing the error code characteristics of packet loss. It is simple to realize and can be used to determine the reason for packet loss in real time. However, COLLIE needs the whole BER packet to be sent back to the sender to analyze the BER characteristics, which leads to a high communication cost.

In this paper, based on the G. Bianchi model [3] method, the calculation formula of frame error rate, collision probability, and throughput in error channel is derived, and the error rate of IEEE 802.11p in Rayleigh channel and different SNR conditions is estimated. By comparing the size of frame error rate and collision rate as the basis of packet loss, a rate adaptive algorithm for VANET environment is designed.

2 Related Work

There is briefly summarization of the distributed coordination function (DCF) as standardized by the 802.11 protocol. For a more complete and detailed presentation, refer to the 802.11 standard [4]. A station with a new packet to transmit monitors the channel activity. If the channel is idle for a period of time equal to a distributed interframe space (DIFS), the station transmits. Otherwise, if the channel is sensed busy (either immediately or during the DIFS), the station persists to monitor the channel until it is measured idle for a DIFS.

G. Bianchi et al. provided a simple, but nevertheless extremely accurate, two-dimensional Markov chain analytical model to compute the 802.11 DCF throughput, in the assumption of finite number of terminals and ideal channel conditions. The ideal

packet collision is caused by at least one of the n-1 nodes in a slot message sending packet data at the same time. Therefore, the probability of frame collision is

$$P_c = 1 - (1 - \tau)^{n-1} \tag{1}$$

Among them, the τ represents the probability of sending packets in any slot, as shown by the following formula (2):

$$\tau = \frac{2(1 - 2P_c)}{(1 - 2P_c)(1 + CW_{\min}) + P_c CW_{\min}(1 - (2P_c)^m)} \tag{2}$$

In this formula, the m represents the maximum back-off stage, and the W_{min} represents the initial competition window. By combining formula (1) and formula (2) can we use the numerical method to find pc and τ under the ideal channel.

Table 1. Modulation code scheme of IEEE 802.11p

Modulation	Code bit rate (Mbps)	Coding rate	Date rate (Mbps)	Date bit per OFDM symbol
BPSK	6	1/2	3	24
BPSK	6	3/4	4.5	36
QPSK	12	1/2	6	48
QPSK	12	3/4	9	72
16-QAM	24	1/2	12	96
16-QAM	24	3/4	18	144
64-QAM	36	2/3	24	192
64-QAM	36	3/4	27	216

Four modulation types and three coding rates are defined in 802.11p protocol. The perceivable data rates are determined by modulation and coding rates used in transmissions. Defined transmission options in IEEE 802.11p are illustrated in Table 1. If M is considered as the MCS vector in which M = {m0, m0. . . . . m7} where m0 = MCS 0 = BPSK rate 1/2 and m7 = MCS 7 = 64-QAM rate 3/4.

Table 2. SNR vs. MCS of IEEE 802.11p

SNR range (dB)	MCS	SNR range (dB)	MCS
<8.5	0	16 ~ 21	4
8.5 ~ 10	1	21 ~ 26	5
10 ~ 14	2	26 ~ 35	6
14 ~ 16	3	>35	7

Rate adaptive algorithm is used to achieve better throughput in the presence of channel errors, so we need to consider how channel errors affect data frames. The higher the signal-to-noise ratio (SNR), the lower the probability of corresponding errors. The corresponding signal-to-noise ratio (SNR) curve of IEEE 802.11p bit error rate (BER) has been calculated in paper [10]. We can see that the higher the transmission rate, the higher the bit error rate in the case of SNR invariant. Next, we will analyze how SNR specifically affects the successful transmission of data.

3 The Proposed Algorithm of PLDRA

When a packet is sent from the upper layer to the MAC layer, the MAC header of 24 bytes is added to the data packet and the frame check sequence (FCS) of 4 bytes is added to the end of the packet which is encapsulated into a frame, called MAC protocol data unit (MPDU). The MPDU is then sent to the physical layer with the physical layer convergence protocol (PLCP) header and preamble. If a frame with data length L is transmitted using modulation mode m, the probability of successful transmission can be expressed as

$$P^m(L) = (1 - p^m_{data}(L)) * (1 - p^m_{ack}) \tag{3}$$

where the $p^m_{data}(L)$ represents the probability of a data frame error of length L, and the p^m_{ack} represents the probability of confirming the frame ACK error. The length of the ACK is much smaller than the data frame, only 14 bytes which have little effect on the transmission error probability and can be ignored. As a result, the expression for the probability of successful transmission can be simplified to

$$P^m(L) \approx (1 - p^m_{data}(L)) \tag{4}$$

Then, we can derive the probability of the data frame error:

$$p^m_{data}(L) = 1 - \left(1 - p^0_e(24)\right) * \left(1 - p^m_e(28 + L)\right) \tag{5}$$

Here $p^0_e(24)$ is the error probability of transmitting the PLCP preamble and header using mode 0. $p^m_e(28 + L)$ is the probability of MPDU. 28 bytes is the MAC header overhead. The bit error rate (BER) can be used to calculate that

$$P^m(L) = 1 - \left(1 - p^m_{ber}\right)^{8L} \tag{6}$$

Here $P^m(L)$ is the bit error rate when using mode m transmission, it increases with the decrease of SNR. The error probability of the received frame can be obtained by substituting formula (6) into formula (5), which is expressed as follows:

$$P_f = 1 - \left(1 - p^0_{ber}\right)^{8*24} * \left(1 - p^m_{ber}\right)^{8*(28+L)} \tag{7}$$

To calculate the error probability (P_f) of the frame, we need to know the SNR value of the frame, the transmission rate (modulation mode m), the length of the data segment L and the corresponding table of BER-SNR. It can be seen that SNR can reflect the channel condition very well, especially when the data frame is longer.

Next, we analyze the DCF performance with the consideration of FER existed. From the previous analysis, we know that the packet transmission failure may be caused by packet collision or it may be caused by channel transmission error. Let P be the packet transmission failure probability, we can get the following formula:

$$P = 1 - (1 - \tau)^{n-1}(1 - P_f) \tag{8}$$

We take Formula (8) into Formula (2) to replace *Pc,* and we can get the packet transmission probability τ under the condition of bit error rate. Next, we derive the system's throughput calculation formula. Combined with the analysis in the previous section, we can see that the channel transmission status can be divided into four types: channel idle, successfully transmitted, collision, and transmission error. The length and probability of these virtual time slots are calculated as follows.

- *Idle*. It means the channel is unused, we define the length of time spent as T_{idle}. As the back-off time is based on slot time. $T_{idle} = SLOT$ and the average probability is $P_{idle} = (1 - \tau)^n$, i.e., the probability that none of the n nodes send data.
- *Successfully transmitted*. It is defined as a successful data exchange. The length defined as T_{succ} under the basic access mechanism, $T_{succ} = DIFS + SIFS + 2DELAY + (H + L)/RATE_{MCS} + ACK/RATE_0$ Where RATEMCS is defined as the corresponding transmission rate for selecting different MCS. $RATE_0$ indicates that the MCS rate is 0. The average probability is $P_{succ} = n\tau(1 - \tau)^{n-1}(1 - P_f)$ which means there is only one sending node, and there is no error in channel transmission.
- *Collision*. It means the packet loss is caused by collision. The length defined as Tcol $T_{coll} = DIFS + (H + L)/RATE_{MCS} + ACK_{timeout}$ Where the $ACK_{timeout}$ is defined as the time when the sender fails to waits for the ACK packet, which can be defined below: $ACK_{timeout} = SIFS + SLOT + ACK/RATE_0 + DELAY$ The average probability of collision is $P_{col} = 1 - (1 - \tau)^n - n\tau(1 - \tau)^{n-1}$.
- *Transmission error*. It means the packet loss is caused by decoding error. It is easy to find that the length of it is the same as collision, i.e., $T_{err} = T_{col}$. The average probability of transmission errors is $P_{err} = n\tau(1 - \tau)^{n-1}P_f$

Based on the above probability of occurrence of various virtual time slots and their length of time spent, we can draw the average length of virtual time slots as

$$T_{average} = P_{idle}T_{idle} + P_{succ}T_{succ} + P_{err}T_{err} + P_{col}T_{col} \tag{9}$$

In the successful slot time, only the data frame load L is valid, so we define the effective data transmission time as $T_{effcient} = L/RATE_{MCS}$ and the throughput S is defined as

$$S = \frac{T_{effcient} P_{succ} RATE_{MCS}}{T_{average}} \tag{10}$$

From the analysis of the previous section, we know that to achieve rate adaptation, the sender needs to know the channel state of the receiver, that is, the statistics of the receiver need to be fed back to the sender and the statistical ACK frame is unable to distinguish between channel error and collision error. The SNR is determined by the power of the transmitter and the distance between the nodes. The collision has no effect on the SNR value. Therefore, using the receiver SNR as the reference variable of the rate adaptive adjustment can effectively measure the channel condition and avoid the impact of link layer collision on the rate adjustment.

We set the quality of service (QoS) of the network to be satisfied by rate adaptation to be BER < 1E-4. According to the graph, we can get the SNR range of each modulation mode and then we can get the throughput SNR diagram by using the band value method according to the formula (10). With the goal of maximizing throughput, a reasonable SNR range for each modulation method in IEEE 802.11p in the VANET environment can be obtained as shown in the Table 2.

We can regard this range as the threshold of MCS switching, but in order to adapt to the complex environment of vehicle networking and avoid the problem of weak signal packet loss caused by the excessive optimization of speed in the literature [6].

The choice of rate does not depend entirely on this threshold. Specifically, the algorithm should make timely speed reduction adjustment to the potentially worse channel environment, so as to ensure smooth communication. For this reason, the algorithm judges the packet loss, and if weak signal packet loss occurs twice in a row, the speed will be reduced. In the case of packet loss, the theoretical packet error rate is obtained according to SNR and compared with the collision probability of the number of surrounding nodes. If the theoretical packet error rate is greater than the collision probability, then the last packet loss is treated as a weak signal packet loss, and if the transmission has not been successful. Then the next transmission will reduce the speed by one file (if it is already minimal, maintain that rate).

The hypothetical sample (pseudocode) in listing below makes it clearer.

```
threshold = [8.5 10 14 16 21 26 35];
snrUp = [threshold inf]+1;
snrDown = [-inf threshold];
increaseMCS = (mean(EstimatedSNR) > snrUp(snrInd));
decreaseMCS = (mean(EstimatedSNR) <= snrDown(snrInd));
snrInd = snrInd+increaseMCS-decreaseMCS;
if  the packet loss in twice
  LossDifference = P_FER > P_COLL
MCS = snrInd - LossDifference;
```

4 Simulation Setting and Results

The simulation is performed on the MATLAB platform using the IEEE 802.11p wireless standard in the WLAN toolbox. Detailed parameters see Table 3.

Table 3. Parameters of the 802.11p and propagation model

Parameter	Value
A slot time	13 μs
A SIFS time	32 μs
A DIFS time	58 μs
CW_{min}	32
CW_{max}	1024
Car number	5\15\25
PSDU Load	1000

From Fig. 1, we can see that the PLDRA frame error rate is about 0.08, better than the ARF about 0.09 and the FARA about 0.5, and the PLDRA throughput is the largest of the three. The PLDRA algorithm can achieve higher accuracy for channel quality prediction and choose the appropriate rate.

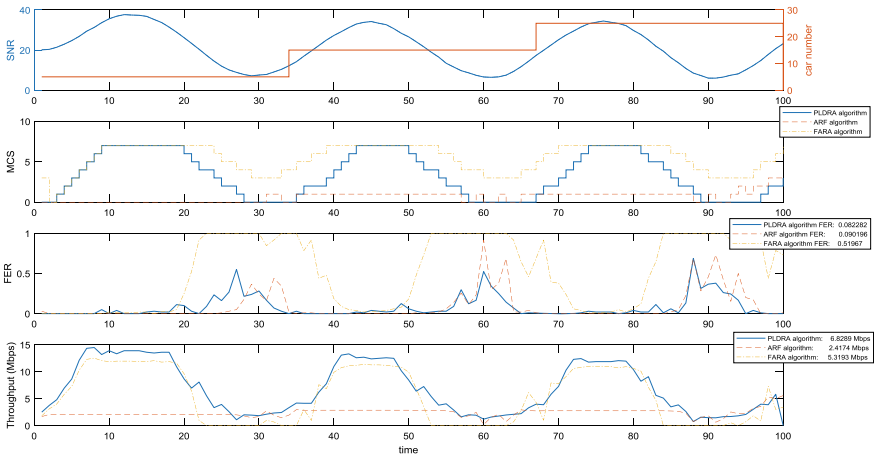


Fig. 1. Simulation results

Figure 2 shows the time-share of each state of the frame transmission, where coll indicates the collision packet error, err indicates the weak signal packet error, idle indicates the back-off waiting time, and succ indicates the time taken by the successful transmission of the frame. From Fig. 2, we can also see that PLDRA has the largest proportion of successful transmission time and the greatest efficiency. The ARF

algorithm takes more time to deal with collisions. The reason is that the algorithm uses lower MCS to transmit, which causes the nodes that compete for the transmission right to occupy longer channel time, which leads to a decrease in the overall system throughput. Since the FARA algorithm is designed for IEEE 802.11a only, it shows a high bit error rate in the VANET environment, so the MAC utilization rate is low.

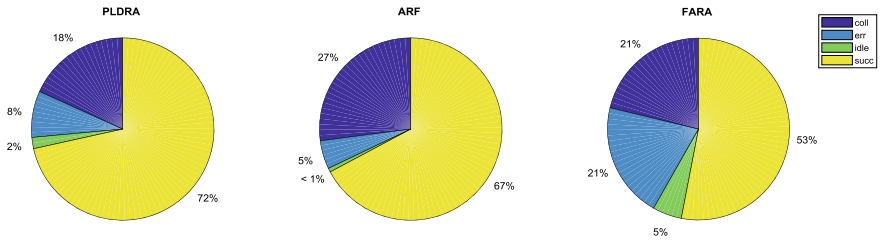


Fig. 2. Pie chart with the percentages of four types of channel transmission

5 Conclusion

A probability-based rate adaptive algorithm with loss differentiation for vehicle network was introduced in this paper to overcome the drawbacks of the existing works in the literature. In our scheme, we proposed a suitable threshold of SNR-MCS switching for IEEE 802.11p and calculated the probability of occurrence of two types of packet loss and used this as the basis for packet loss judgment. Thus, the performance of throughput and FER is kept in reasonably good condition. The working principle of our algorithm has been evaluated through computer simulation, and the obtained results have proven its efficiency.

Acknowledgements. Shubin Wang (wangshubin@imu.edu.cn) is the correspondent author. This work was supported by the National Natural Science Foundation of China (61761034, 61261020); the Natural Science Foundation of Inner Mongolia, China (2016MS0616); and the Enhancing Comprehensive Strength Foundation of Inner Mongolia University (No. 10000-16010109-57).

References

1. Alasmary, W., Zhuang, W.: Mobility impact in IEEE 802.11p infrastructure less vehicular networks. Ad Hoc Networks **10**(2), 222–230 (2012)
2. Atallah, R., Khabbaz, M., Assi, C.: Multi-hop vehicle-to-infrastructure communications: a feasibility study, modelling and performance analysis. IEEE Trans. Vehicular Technol. **PP** (99), 1–1 (2017)
3. Bianchi, G.: Performance analysis of the IEEE 802.11 distributed coordination function. IEEE J Select. Areas Commun. **18**(3), 535–547 (2000)
4. IEEE. Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE Std (1997)

5. Kamerman, A., Monteban, L.: WaveLAN II: a high-performance wireless LAN for the unlicensed band. Bell. Labs Techn. J. **2**(3), 118–133 (1997)
6. Khan, M., Veitch, D.: Isolating physical per for smart rate selection in 802.11. In: Proceedings of INFOCOM 2009. [S. l.]: IEEE Press, pp. 1080–1088 (2009)
7. Kim, S., Verma, L., Choi, S., et al.: Collision-aware rate adaptation in multi-rate WLANs: design and implementation. Comput. Netw. **54**(17), 3011–3030 (2010)
8. Rahul, H., Edalat, F., Katabi, D., et al.: Frequency-aware rate adaptation and MAC protocols. In: Proceedings of MOBICOM, pp. 193–204 (2009)
9. Rayanchu, S., Mishra, A., Agrawal, D., et al.: Diagnosing wireless packet losses in 802.11: separating collision from weak signal. In: Proceedings of INFOCOM 2008. [S. l.]: IEEE Press, pp. 735–743 (2008)
10. Sassi, A., Charfi, F., Hillali, Y.E., et al.: OFDM transmission performance evaluation in V2X communication. Int. J. Comput. Sci. Issues **9**(2) (2012)

Optimal Design on UAV-Enabled Downlink Wireless Information and Energy Transfer

Yabin Lai, Yue Ling Che, Sheng Luo, Jie Ouyang, and Kaishun Wu(✉)

Shenzhen University, Shenzhen, China
{laiyabin2016,ouyangjie2016}@email.szu.edu.cn,
{yuelingche,sluo,wu}@szu.edu.cn

Abstract. This paper investigates wireless information transfer (WIT) and wireless energy transfer (WET) in an unmanned aerial vehicle (UAV) enabled downlink wireless communication system, where both the UAV and the ground terminal (GT) are battery-empowered. When the UAV transmits energy to the GT, the GT harvests energy and stores in its battery, and when the UAV transmits information to the GT, the GT uses the stored energy from the battery to receive and decode the information. We formulate the UAV's optimal WIT and WET design as a constrained Markov decision process (MDP) problem over a finite time horizon. Due to the high complexity to solve this problem, we then propose a suboptimal policy by studying the optimal policy structure. Numerical results show that the proposed suboptimal policy can achieve close-to-optimal performance. We also show the impact of the UAV's altitude on the throughput performance.

Keywords: Unmanned aerial vehicle (UAV) · Wireless information transfer (WIT) · Wireless energy transfer (WET) · Markov decision process (MDP) · Battery management

1 Introduction

Recently, UAV-enabled/assisted wireless information transfer (WIT) has attracted substantial attentions, due to their flexible deployment, reconfigurable operations and the large chance to have strong line-of-sight (LoS) communication channels with the ground terminals (GTs). For example, in [1], the UAVs can be fast deployed as aerial base stations if the terrestrial communication fails during natural disasters or under malicious attacks. In [2], an UAV-enabled mobile relaying system, where the UAV relays information from the source to the destination, was proposed. The source/relay transmit power as well as the UAV trajectory were jointly optimized to maximize the system throughput.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 490–498, 2019.
https://doi.org/10.1007/978-981-13-6264-4_59

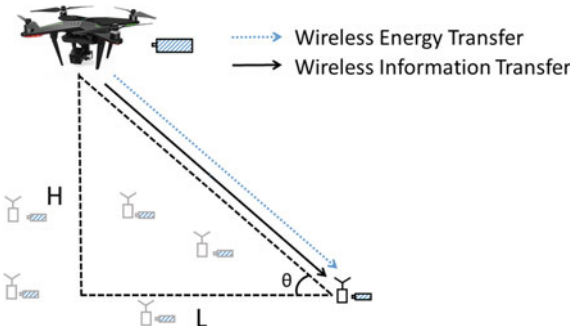


Fig. 1. System model.

Moreover, the UAV-enabled wireless power transfer (WPT) has been proposed in [3] where the UAV transmits wireless energy over the radio frequency (RF) spectrum to the GTs. As compared to the traditional WPT system with power beacons deployed on ground, the flexible deployment of the UAV can assure short-distance WPT to the GTs [4]. Further, [5] considered an UAV-enabled wireless powered communication network (WPCN), where the GTs use the harvested energy from the UAV to support their WITs to the UAV.

However, the UAVs are normally battery-empowered with limited energy. Such an energy sustainability issue of the UAVs has not been discussed in the existing studies for efficient WIT/WET, to our best knowledge. In this paper, we consider that both the UAV and the GTs are battery-empowered, and study the effects of efficient battery management to improve the system throughput. As shown in Fig. 1, we consider an UAV-enabled downlink communication system. When the UAV transmits energy to the GTs, the GTs harvest energy from the UAV's RF signals; when the UAV transmits information to the GTs, the GTs consume their battery to receive and decode information. The UAV's optimal WIT and WET design can be formulated as a constrained Markov decision process (MDP) problem over a finite time horizon. Due to the high complexity to solve this problem, we propose a suboptimal policy by studying the optimal policy structure. Numerical results show that the proposed suboptimal policy can achieve close-to-optimal performance. We also show the impact of the UAV's altitude on the throughput.

2 System Model

We consider an UAV-enabled downlink wireless communication system. As shown in Fig. 1, an UAV is dispatched to serve a set of GT on ground. The UAV is installed with a rechargeable battery and exploits energy from the battery to support all its operations. For saving the UAV's limited energy, we assume that the UAV hovers at a certain location above the set of GTs. We assume that each GT is also installed with a battery, such that it can store energy harvested from the UAV's WETs. We denote the battery capacity of each GT as B_{max} with

$0 < B_{max} < \infty$. All the GTs access the spectrum in a time-division multiple-access (TDMA) manner. We then focus on the downlink communication from the UAV to a single GT.[1] We assume that the UAV exploits E_p (Wh) amounts of energy from its battery, where $E_p > B_{max}$, to support its WITs and WETs to the GT over a finite time horizon within total T slots.[2] As shown in Fig. 1, we denote the horizontal distance between the UAV and the GT as L. At the beginning of each slot $t \in \{1, ..., T\}$, the UAV chooses its transmission action $a(t)$ among keeping silent with $a(t) = 0$, WET with $a(t) = 1$, and WIT with $a(t) = 2$. We assume that the UAV adopts a fixed transmit power $P_f > 0$ for both WET and WIT. Thus, according to the transmission action $a(t)$, the UAV's transmit power is given as

$$P(t) = \begin{cases} 0, a(t) = 0, \\ P_f, a(t) = 1 \text{ or } a(t) = 2. \end{cases} \tag{1}$$

For the air-to-ground (AtG) channel from the UAV to the GT, the GT generally receives three groups of signals, which are the LoS signals, non-line-of-sight (NLoS) signals and multiple reflected signal components [6]. As in [6], we consider a probability-based AtG channel model to include all three groups of signals. Specifically, as shown in Fig. 1, supposing that the UAV is elevated at the altitude of H, the LoS probability is given as

$$p_L = \frac{1}{1 + a \exp(-b(\theta - a))}, \tag{2}$$

where a and b are constant values depending on the network environment, and $\theta = \frac{180}{\pi} arctan \frac{H}{L}$ is the elevation angle. The NLoS probability is thus $p_N = 1 - p_L$. We assume Nakagami-m fading for both the LoS connection and the NLoS connection. Then the AtG channel state in slot t is given by

$$\gamma(t) = \begin{cases} h_L(t)(\sqrt{L^2 + H^2})^{-\alpha_L}, & \text{if LoS}, \\ h_N(t)(\sqrt{L^2 + H^2})^{-\alpha_N}, & \text{if NLoS}, \end{cases} \tag{3}$$

where α_L and α_N are the path loss exponents of the LoS connection and NLoS connection, respectively. In (3), $h_L(t)$ for LoS connection and $h_N(t)$ for NLoS connection are the normalized Gamma random variables, corresponding to the Nakagami-m fading gain with the parameters m_L and m_N, respectively. We assume $h_L(t)$'s and $h_N(t)$'s, $t \in \{1, ..., T\}$, are mutually independent. As a result, the received signal power strength at the GT in each slot t is obtained as $\eta P(t)\gamma(t)$, where $0 < \eta < 1$ is the energy harvesting efficiency at the GT.

[1] Results of the single-GT scenario in this paper can be extended to the multi-GT scenario and will be studied in our future work.

[2] Besides supporting the downlink transmissions, we assume that the UAV also consumes a constant amount of energy to support all other operations (e.g., propulsion in air) during the T slots, which is thus not considered in the UAV transmission policy design.

Denote $E_r(t)$ as the residual energy available at the UAV in slot t. We thus have $E_r(1) = E_p$. In this paper, we formulate the UAV's transmission decisions over time as a constrained MDP, where the UAV determines its transmission action $a(t)$ in each slot t based on its residual energy level $E_r(t)$, the GT's battery level $B(t)$, as well as the AtG channel state $\gamma(t)$. Details of the problem formulation are given in the next section.

3 Problem Formulation

In this section, we formulate the UAV's optimal WIT/WET transmission decision as a constrained Markov decision process (MDP) problem over T slots, and then study the optimal policy structure.

We first introduce the basic elements in the MDP problem, which includes the system state space, action space, system state update, and the instantaneous reward. Specifically, the state space of the MDP is given as $\mathbb{S} = \{S(t) = (\gamma(t), B(t), E_r(t)) : \gamma(t) \in [0, +\infty), B(t) \in [0, B_{max}], E_r(t) \in [0, E_p]\}$. The system state space is generally infinitely large. The action space is given by $\mathbb{A} = \{a(t) : a(t) \in \{0, 1, 2\}\}$.

According to the UAV's transmission action $a(t)$ in slot t, $t \in \{1, ..., T-1\}$, the GT's battery level at the beginning of slot $t+1$ is obtained as

$$B(t+1) = \begin{cases} B(t), & a(t) = 0 \\ min\{B(t) + \eta P(t)\gamma(t), B_{max}\}, & a(t) = 1 \\ B(t) - E_d, & a(t) = 2 \end{cases} \tag{4}$$

where as compared to $B(t)$, $B(t+1)$ is increased via GT's energy harvesting when the UAV selects WIT, is reduced for information receiving and decoding when the UAV selects WET, and is unchanged when the UAV is silent. Similarly, the UAV's residual energy level at the beginning of slot $t+1$ is obtained as

$$E_r(t+1) = \begin{cases} E_r(t), & a(t) = 0 \\ E_r(t) - P_f, & a(t) = 1 \text{ or } a(t) = 2 \end{cases} \tag{5}$$

where as compared to $E(t)$, $E(t+1)$ is reduced by P_f for WIT or WET, and is unchanged when silent. Hence, both $B(t)$ and $E_r(t)$ are only depending on that in the previous slot, respectively. Moreover, as mentioned in Sect. 2, $\gamma(t)$'s are mutually independent. As a result, the Markovian property is satisfied. Based on the system state $S(t)$ and the UAV's transmission action $a(t)$, the instantaneous reward is obtained as

$$R_t(S(t), a(t)) = \log_2\left(1 + \frac{\eta P(t)\gamma(t)}{P_0}\right) I(B(t) \geq E_d, E_r(t) \geq P_f, a(t) = 2), \tag{6}$$

where P_0 is the noise strength, and $I(x_1, x_2, x_3)$ is the indicator function with $I(x_1, x_2, x_3) = 1$ if x_1, x_2 and x_3 are all true, and $I(x_1, x_2, x_3) = 0$, otherwise.

Denote the UAV's policy as $\pi = \{J(t)\}$, $t \in \{1, ..., T\}$, where $J(t)$ in slot t maps the system state $S(t)$ to the UAV's action decision $a(t)$. Our goal is to find

the optimal UAV policy π^* such that the UAV's sum reward over the T slots is maximized. The MDP problem is therefore formulated as

$$(P1): \quad \max_{\pi} \sum_{t=1}^{T} \mathbb{E}\left[R_t(S(t), a(t)\right] \tag{7}$$

$$s.t. \quad E_r(T) - P(T) \geq 0, \tag{8}$$

where the constraint in (8) ensures that the residual energy at the UAV is always nonnegative. According to [7], the optimal π^* can be found by using a backward induction method. The value function is given as

$$V_t(S(t)) = \begin{cases} \max_{a(T)} R_T(S(T), a(T)), & if \quad t = T \\ \max_{a(t)} R_t(S(t), a(t)) + V_{t+1}(S(t+1)|S(t)), & else. \end{cases} \tag{9}$$

However, due to the generally infinitely large system state space, it is of high complexity to calculate $V_t(S(t))$ and thus the optimal π^* [7]. In the following, by finding the optimal $a(t)^*$ in the last two slots, we study the optimal policy structure, based on which a suboptimal policy is proposed in the next section.

Table 1. Optimal $a(T-1)^*$

$B(t)$	$E_r(t)$		
	$E_r(t) < P_f$	$P_f \leq E_r(t) < 2P_f$	$E_r(t) \geq 2P_f$
$B(t) < E_d$	$a(t)^* = 0$	$a(t)^* = 0$	$a(t)^* = 1$
$E_d \leq B(t) < 2E_d$	$a(t)^* = 0$	(11)	(12)
$B(t) \geq 2E_d$	$a(t)^* = 0$	(11)	$a(t)^* = 2$

First, for the last slot with $t = T$, problem (P1) is reduced to maximize $R_T(S(T), a(T))$. It is thus easy to find the optimal action $a(T)^*$ as

$$a(T)^* = \begin{cases} 2, & if \quad B(T) \geq E_d, E_r(T) \geq P_f \\ 0, & else. \end{cases} \tag{10}$$

Next, for the slot $t = T-1$, problem (P1) is reduced to maximize $R_{T-1}(S(T-1), a(T-1)) + \mathbb{E}\left[R_T(S(T), a(T))\right]$. Using a method similar to that in [8], we obtain $a(T-1)^*$ as shown in Table 1, where when $P_f \leq E_r(T-1) < 2P_f$, $E_d \leq B(T-1) < 2E_d$, or $P_f \leq E_r(T-1) < 2P_f$, $B(T-1) \geq 2E_d$, we have

$$a(T-1)^* = \begin{cases} 2, & if \quad \log_2\left(1 + \frac{\eta P_f \gamma(T-1)}{P_0}\right) > \mathbb{E}\left[\log_2\left(1 + \frac{\eta P_f \gamma(T)}{P_0}\right)\right] \\ 0, & else. \end{cases} \tag{11}$$

When $E_r(T-1) \geq 2P_f$ and $E_d \leq B(T-1) < 2E_d$, we have

$$a(T-1)^* = \begin{cases} 2, & \log_2\left(1 + \frac{\eta P_f \gamma(T-1)}{P_0}\right) > \mathbb{E}\left[\log_2\left(1 + \frac{\eta P_f \gamma(T)}{P_0}\right)\right] \\ 1, & else. \end{cases} \tag{12}$$

It is observed from (10) and Table 1 that the UAV's transmission action is strongly related to the value of $B(t)$ and $E_r(t)$. If the UAV's residual energy is lower than the required transmit power P_f, it selects to be silent in both (10) and Table 2. If both $B(t)$ and $E_r(t)$ are very high to completely satisfy the GT's and the UAV's energy requirement, the UAV always selects WIT to improve throughput. Moreover, if $B(t) < E_d$, the UAV selects being silent for saving energy or WET for harvesting energy, according to whether the UAV has a large amount of residual energy. In all other cases, $a(t)^*$ is related to the channel state $\gamma(t)$. Based on the above observations, we propose an efficient suboptimal policy in the next section.

4 Suboptimal Policy

This section proposes the suboptimal policy according to the observation of the optimal policy structure in Sect. 3. First, similar to the action decision in Table 1, to ensure that the constraint in problem (P1) holds, we assume that when $E_r(t) < P_f$, the UAV keeps silent from slot t to slot T. Next, we only consider the case with $E_r(t) \geq P_f$ in the following.

Specifically, the battery level of the GT is divided into three regimes, i.e., low GT battery regime with $B(t) < E_d$, medium GT battery regime with $E_d \leq B(t) < (1+T-t)E_d$ and high GT battery regime with $B(t) \geq (1+T-t)E_d$, respectively. Similarly, the UAV's residual energy level is also divided into three regimes, i.e., low UAV battery regime with $P_f \leq E_r(\mathrm{t}) < 2P_f$, medium UAV battery regime with $2P_f \leq E_r(\mathrm{t}) < (1+T-t)P_f$ and high UAV battery regime with $E_r(\mathrm{t}) \geq (1+T-t)P_f$, respectively.

Table 2. Suboptimal $a(t)^\star$ for $E_r(t) \geq P_f$

$B(t)$ $E_r(t)$	$E_r(t)$		
	Low	Medium	High
Low	$a(t)^\star = 0$	(14)	$a(t)^\star = 1$
Medium	(14)	(14)	(14)
High	(14)	(14)	$a(t)^\star = 2$

The suboptimal policy is given in Table 2, where the suboptimal $a(t)^\star$ for the cases with (1) low $B(t)$ and low $E_r(t)$, (2) low $B(t)$ and high $E_r(t)$, and (3) high $B(t)$ and high $E_r(t)$ can be easily obtained based on the optimal ones in Table 1. For all other cases, we calculate the suboptimal $a(t)^\star$ by approximately calculating the value function in (9). To be specific, we define an estimated value function in slot $t \in \{1, ..., T-1\}$ to approximate $V_t(S(t))$ as

$$Q_t(S(t), a(t)) \triangleq R_t(S(t), a(t)) + F_t(B(t+1), E_r(t+1)). \tag{13}$$

In (13), $F_t(B(t+1), E_r(t+1)) \triangleq n_m \cdot \log_2\left(1 + \frac{\eta P_f E(\gamma)}{P_0}\right)$ is the estimated throughput accumulated from slot $t+1$ to slot T, where n_m is the estimated number of

WITs from slot $t+1$ to slot T, and $E(\gamma)$ is the channel state with $E(\gamma) = p_L/m_L(\sqrt{L^2+H^2})^{-\alpha_L} + p_N/m_N(\sqrt{L^2+H^2})^{-\alpha_N}$. When the UAV's residual energy is low and the GT's battery level is medium or high, $F_t(B(t+1), E_r(t+1))$ is given by $\log_2\left(1+\frac{\eta P_f E(\gamma)}{P_0}\right)$ if $a(t)=0$, and given by 0 if $a(t)=2$. When the UAV's residual energy is medium, the estimated number of WETs of the UAV is given by $n_c = \arg\min_n \left|\left\lfloor \frac{E_r(t+1)}{P_f}\right\rfloor - n - \left\lfloor \frac{B(t+1)+nP_fE\{\gamma\}}{E_d}\right\rfloor\right|$, where $\left\lfloor \frac{E_r(t+1)}{P_f}\right\rfloor - n$ denotes the number of WITs that UAV operates after n times' WETs, and $\left\lfloor \frac{B(t+1)+nP_fE\{\gamma\}}{E_d}\right\rfloor$ denotes the number of times that the GT can receive and decode information after being charged by the UAV for n times during slot $t+1$ to T. The future reward is decided by the number of future WITs which is given by $n_m = min\left\{\left\lfloor \frac{E_r(t+1)}{P_f}\right\rfloor - n_c, \left\lfloor \frac{B(t+1)+n_cP_fE\{\gamma\}}{E_d}\right\rfloor\right\}$. When the UAV's energy is high and the GT's battery level is medium, the future number of WETs for the UAV is given by $n_c = \arg\min_n \left|\lceil T-t\rceil - n - \lfloor \frac{B(t+1)+nP_fE\{\gamma\}}{E_d}\rfloor\right|$, where $\lceil T-t\rceil - n$ denotes the future slot number for the UAV to do WIT after n times' WET, and $\lfloor \frac{B(t+1)+nP_fE\{\gamma\}}{E_d}\rfloor$ denotes the number of times that the GT can receive and decode information after being charged by the UAV for n times during slot $t+1$ to T. Under this situation, the future number of WITs is given by $n_m = min\left\{\lceil T-t\rceil - n_c, \lfloor \frac{B(t+1)+n_cP_fE\{\gamma\}}{E_d}\rfloor\right\}$. Finally, the action in slot t is given by

$$a(t)^\star = \arg\max_{a_t} Q_t(S(t), a_t). \tag{14}$$

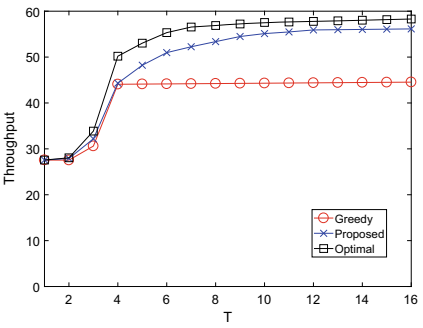


Fig. 2. Throughput comparison.

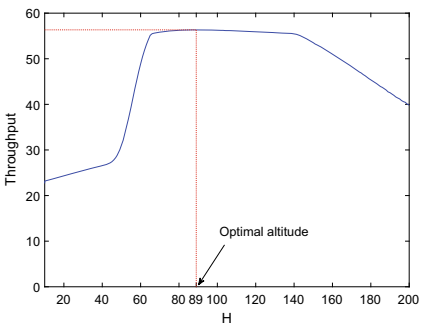


Fig. 3. Impact of the UAV's altitude.

5 Simulation Results

Numerical results are given in this section. We set $L = 200$ m, $P_f = 100$ mW, $P_0 = -100$ dBm, $m_L = 3$, $m_N = 2$, $a = 8.5$, $b = 0.33$, $E_d = 4\ \mu W \cdot s$, $E_p = 40$

$mW \cdot s$ and $B(1) = 4\ \mu W \cdot s$. For comparison, we consider the optimal policy and a greedy policy. For the greedy policy, the UAV always selects to transmit as long as its residual energy is not lower than P_f. In this case, the UAV selects WIT if the GT's battery $B(t) \geq E_d$, and WET otherwise. Under each policy and for each T, we run 1000 times and calculate the average throughput. It is observed from Fig. 2 that when T is not larger than 5, the throughputs of the proposed policy and the greedy policy are almost the same, because the UAV has enough energy to transmit when T is small. As T approaching to 16, the proposed policy shows significantly improved performance as compared to the greedy policy. When $T = 16$, the throughput of the proposed policy is 26.05% higher than that of the greedy policy while only 3.84% less than that of the optimal policy. Thus, the proposed policy shows close-to-optimal performance when T increases.

We also show the impact of the UAV's altitude on the throughput when $T = 16$ in Fig. 3 under the proposed policy. It is observed that the system throughput first increases then decreases over the UAV altitude H, and achieves maximal value when $H = 89\,\text{m}$.

6 Conclusion

This paper studied the optimal WIT and WET design over time for an UAV-enabled downlink wireless communication system. We formulated the problem as a constrained MDP problem. Due to the complexity to find the optimal policy, we proposed an efficient suboptimal policy. Simulation results verified the high performance of the proposed policy.

Acknowledgements. This work was supported in part by grants from China NSFC (61601308, 61872248), Guangdong Provincial Science and Technology Development Special Fund project (2017A010101033, 2017A030312008), Shenzhen Science and Technology Funding (JCYJ20170818093658379, JCYJ20170302140946299, JCYJ20170412110753954), Tencent "Rhinoceros Birds"-Scientific Research Foundation for Young Teachers of Shenzhen University,Guangdong Talent Project 2015TX01X111 and GDUPS (2015).

References

1. Merwaday, A., Guvenc, I.: UAV assisted heterogeneous networks for public safety communications. In: Proceedings IEEE Wireless Communications Network Conference, pp. 329–334, Mar. 2015
2. Zeng, Y., Zhang, R., Lim, T.J.: Throughput maximization for UAV-enabled mobile relaying systems. IEEE Trans. Commun. **64**(12), 4983–4996 (2016)
3. Xu, J., Zeng, Y., Zhang, R.: UAV-enabled wireless power transfer: trajectory design and energy optimization. IEEE Trans. Wirel. Commun., forthcoming. Available online at https://arxiv.org/abs/1709.07590
4. Hua, M., Li, C., Huang, Y., Yang, L.: Throughput maximization for UAV-enabled wireless power transfer in relaying system. In: Proceedings IEEE WCSP

5. Che, Y.L., Duan, L., Zhang, R.: Spatial throughput maximization in wireless powered communication networks. IEEE J. Sel. Areas Commun. **33**(8), 1534–1548 (2015)
6. Zhu, Y., Zheng, G., Fitch, M.: Secrecy rate analysis of UAV-enabled mm Wave networks using Matern hardcore point processes. IEEE J. Sel. Areas Commun. forthcoming
7. Puterman, M.L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley-Interscience (1994)
8. Che, Y.L., Zhang, R., Gong, Y.: On design of opportunistic spectrum access in the presence of reactive primary users. IEEE Trans. Commun. **61**(7), 2678–2691 (2013)

CNSA: A Candidate Nodes Selection Algorithm for VANETs

Leilei Qi[1,2(✉)]

[1] College of Information Science and Technology, Donghua University, Shanghai 201620, China
jeremyqilei@163.com
[2] Engineering Research Center of Digitized Textile and Apparel Technology, Ministry of Education, Shanghai 201620, China

Abstract. In order to improve the efficiency of data transmission between nodes in network communication, certain problems such as the instability of nodes and the routing performance, especially the selection of the candidate nodes, are needed to be resolved. In this paper, we considered many factors according to our vehicle motion model to enhance the efficiency of BNSR which is the supplement of BMFR routing protocol. The algorithm that we present is CNSA: A Candidate Nodes Selection Algorithm for VANETs, which constructs a value composed of the relative distance, lifetime of packet, forward angle, and node density for choosing the best candidate to transmit packet. In the end, our simulation is accomplished by using NS2 simulator. As we considered more factors, the final consequence indicates that our algorithm has more advantages than the BNSR protocol, especially in terms of end-to-end delay.

Keywords: BNSR · BMFR · CNSA · NS2 · End-to-end delay

1 Introduction

In recent years, VANET is becoming a very significant field in modern society. As a result of the utilization of vehicles and more frequent availability of information, the drivers are able to drive more safely to the destination. However, VANET has some limitations such as a quickly changing network topological structure and shorter routing time. Therefore, there is an ever-increasing demand to design an efficient and more reliable algorithm to transmit information [1]. Vehicle-to-vehicle (V2V) and vehicle-to-road side unit (V2R) communication are two ways in the high dynamic network to guarantee that the messages can be safely transmitted to other vehicles. As a result of high cost of base station construction, the use of V2V scenario is necessary when there is no RSU along the road [2]. Dedicated short-range communication (DSRC) is presented to support communication applications. It offers more than 27 mbps data rate by working on 5.9 GHz radio signal frequency [3]. The major work of routing is to find the optimal path when the source node wants to establish a connection with the

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 499–506, 2019.
https://doi.org/10.1007/978-981-13-6264-4_60

destination node. Because of frequently changing network topology in VANETs, building a stable link is quite important when the destination node is out of the transmission range of the source node [4].

In the past years, several position-based routing protocols used in VANETs have been proposed such as GPSR, MFR, BFME, BNSR, etc. One of typical position-based routing protocol is GPSR [2] protocol that takes geographic position information into consideration. The Border Node Preferred Social Ranking (BNSR) [5] Based Routing Protocol for VANETs is the expansion of the BMFR [6] routing protocol.

In this paper, a position-based routing is proposed for VANETs which combines the advantages of B-MFR and BNSR. According to the graphics and the environment of vehicles, the next better candidate node is selected to transfer the data to the final node. A more stable link is created because of the better delivery of packets and shorter delay of end-to-end.

The rest of paper is organized as follows: we construct our main vehicle motion model in Sect. 2. In Sect. 3, we present the additional factor and purpose protocol. Section 4 shows the performance of our algorithm and analysis. The final section is about conclusion.

2 System Model

2.1 Model Frame

In Fig. 1, we assumed a vehicle motion model. In this model, every node can send and receive "Hello Message". The "Hello Message packet" format is revised as in Table 1. Once one node wants to communicate with other nodes, it will be regarded as the source node and the "Hello Message" that include information about the direction and the position of vehicles will be selected and be provided to it. We also define destination node as D. As we know in the vehicular ad hoc networks, every single node has its transmission range and the nodes (red nodes) which lie in this area are called the neighbor nodes. Meanwhile, some (green nodes) of the neighbor nodes will be chosen to be the candidate node to rebroadcast the packet. As the model shows, the D is out of transmission range of S, so they cannot communicate by one hop. So the C_1 will be selected the first forwarding hop node, because it has the best w_i we proposed in this paper transfer the data packets to the final node D. After that, the C_2, C_3 will be chosen successively to be the candidate nodes.

Table 1. The hello message packet

Format of hello message packet					
Src Id	Dest Id	Next-hop Id	Dest position	Time stamp	W_{c_i}

2.2 Description of Parameters

By using GPS system, it can provide the information such as the coordinates of vehicles and velocities of vehicles. We assume that node S is the source node, and node C is the candidate node to receive for the packet. And node C is in the transmission range of node S. The position of node S and node C are, respectively, S (x_s, y_s), and C (x_c, y_c). The speed of node S and node C are, respectively, $\overrightarrow{v_s}$, $\overrightarrow{v_c}$. The distance between node S and node C is defined as D_{sc}. The $\overrightarrow{v_{cs}}$ is the relative velocity when the node S as the reference. After time T_c, we assume that node C will be out of the maximum transmission range R of node S and move to C′. The relative distance of C and C′ is defined as $D_{cc'}$. The motion model is shown in Fig. 2.

According to the coordinates of nodes, the D_{sc} can be calculated by

$$D_{sc} = \sqrt{(x_c - x_s)^2 + (y_c - y_s)^2}. \tag{1}$$

Combining with R, θ, d_{sc}, and cosine formula, the R can be calculated by

$$R^2 = [D_{cc'} \cos(\pi - \theta) + D_{sc}^2] + (D_{cc'} \sin\theta)^2 \tag{2}$$

so the $D_{cc'}$ can be easily get by

$$D_{cc'} = \sqrt{R^2 - D_{sc} \sin^2\theta} + D_{sc} \cos\theta. \tag{3}$$

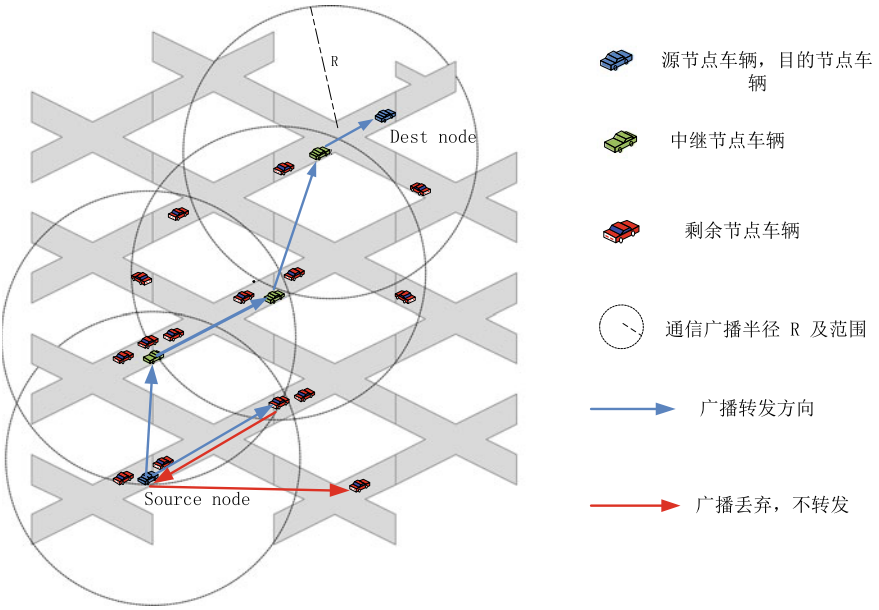


Fig. 1. Vehicle motion model

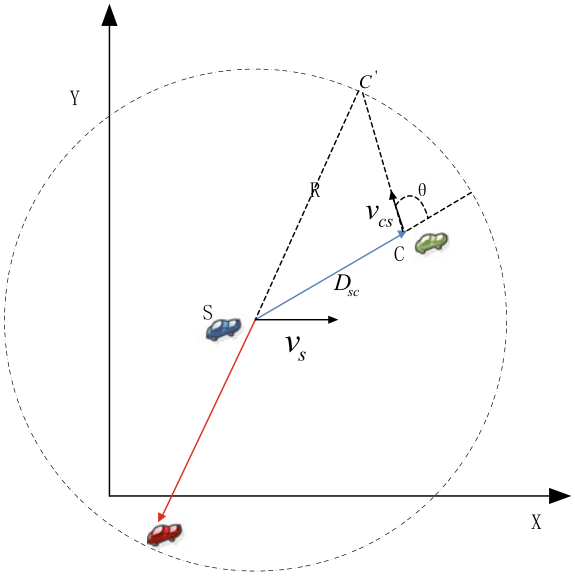


Fig. 2. Candidate node selection based on distance

The relative velocity of node S and node C is $\overrightarrow{v}$

$$\overrightarrow{v_{cs}} = \overrightarrow{v_s} - \overrightarrow{v_c}. \tag{4}$$

In the end, according to Eqs. (3) and (4), the time T_c can be calculated by

$$T_c = \frac{D_{cc'}}{|\overrightarrow{v_{cs}}|} = \frac{\sqrt{R^2 - D_{sc}\sin^2\theta} + D_{sc}\cos\theta}{|\overrightarrow{v_s} - \overrightarrow{v_c}|}. \tag{5}$$

In Fig. 3, we assume that Node D is the destination node, and D_{cd} is the distance of node C and node D. D $(x_d,\ y_d)$ is given to be the coordinate of node D

$$D_{cd} = \sqrt{(x_d - x_c)^2 + (y_d - y_c)^2}. \tag{6}$$

And we also define α is the angle between the movement direction of C and the next-hop candidate node as Fig. 3 shown.

In Fig. 4, we take another factor into consideration which is defined as ρ_{c_i}. N_{c_i} denotes the total number of neighbor nodes around c_i in the transmission R. The ρ_{c_i} is defined as

$$\rho_{c_i} = \begin{cases} \frac{N_{c_i}}{\pi R^2} & D_{c_i d} \neq 0 \\ 0 & D_{c_i d} = 0 \end{cases} \tag{7}$$

as shown in Fig. 4, the packet is always sent to areas with high density in case message pause.

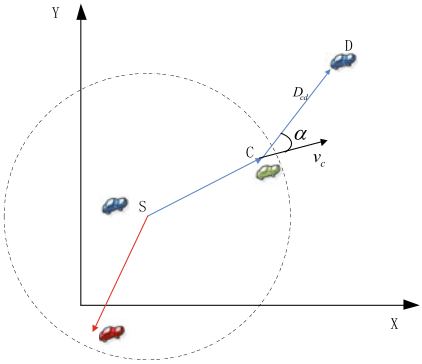


Fig. 3. Candidate node selection based on angle

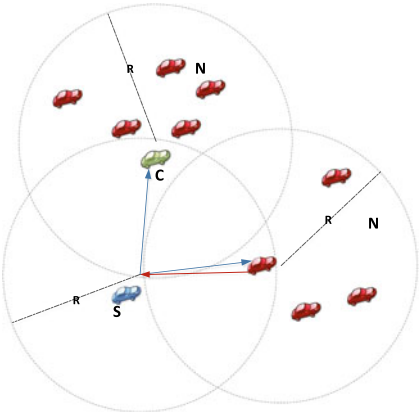


Fig. 4. Candidate node selection based on relay nodes' number

3 Candidate Nodes Selection Algorithm

As mentioned above, we have discussed some factors for choosing next-hop candidate node and making the link more stable. We design final value W_{c_i} based on the model as follows:

$$W_{c_i} = \beta_1 \frac{1}{T_{c_i}} + \beta_2 D_{c_i d} + \beta_3 \sin \alpha_{c_i d} + \beta_4 \frac{1}{\rho_{c_i}}, \tag{8}$$

where $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1$, T_{c_i}, $D_{c_i d}$, and $\sin \alpha_{c_i d}$ are according to Eqs. (5), (6), (7). S denotes the node of sending packet, C denotes the next-hop candidate node, and D denotes the final node of receiving packet. The next-hop candidate node selection algorithm is shown in **Algorithm 1**.

4 Simulation and Analysis

4.1 Simulation Environment

TABLE II

To evaluate the performance of our algorithm, we set β_1 as 0.3, β_2 as 0.2, β_3 as 0.3, and β_4 as 0.2. We utilized the network simulator NS2 and also simulated three algorithms BMFR, BNSR, and GPSR as theoretical references and performance comparison. Table 2 is the simulation environment with detailed settings. We mainly simulate and analyze the performance of packet delivery ratio (PDR) and end-to-end delay.

Table 2. NS2 simulation parameters settings

Simulation time	600 s
Model area	2400 m * 2400 m
No. of vehicles	10, 30, 50, 70, 90, 130
Speed of vehicles/km/h	15,30,45,60,75,90
Type of MAC	IEEE 802.11b
Broadcast range	250 m
Size of packet	512 bytes

Algorithm 1 Process for selecting next candidate node

{Variable C, Z, β_1, β_2, β_3, β_4, W_{c_i}, ρ_{c_i}, T_{c_i}, c_i, $(0 \leqslant \mathrm{i} < Z)$, Min}
{The S wants to communicate with D}
for Every packet broadcasted by S **do**
  if Node D is not out of the transmission R of node S **then**
    Node S transmits packet to Node D directly.
  else
    if There are Z candidate nodes **then**
      calculate every W_{c_i} of C=$\{c_1, c_2 \ldots c_z\}$
      for (i=0, i< Z, i++) **do** State $\{W_{c_i} = \beta_1 \frac{1}{T_{c_i}} + \beta_2 D_{c_i d} + \beta_3 \sin \alpha_{c_i d} + \beta_4 \frac{1}{\rho_{c_i}}\}$
        if (W_{c_i} < Min) **then**
          {Min = W_{c_i}}
          c_i is selected to be next candidate node.
        end if
      end for
    else
      the sender keeps going on road.
    end if
  end if
end for
{Above conditions are true and choose the minimum value W_{c_i} for next candidate node}
{The node D receive data packet}

4.2 Simulation Results and Performance Analysis

End-to-End Delay Performance Analysis. Figure 5 shows the influence of vehicle nodes on transmission delay. In the simulation environment of Table 2. As we see CNSA is obviously better than the performance of GPSR, BMFR, BNSR according to the W_{c_i} set by CNSA for selecting next better candidate node. In the situation where the vehicle nodes are below than 30, the nodes are relatively sparse by comparison of 2400 m * 2400 m vehicle region. Route discovery takes more time to be found slowly and we find that the average end-to-end delay is

a little high. With the increasing number of nodes up to 70, the route discovery can be found quickly and the average end-to-end delay is getting lower.

Figure 6 illustrates the relationship between E2E Delay and the speed of vehicles. In the simulation environment of Table 2. There is no big difference between the four protocols when the speed of the nodes is slower, and the link of data is relatively stable. As the speed of vehicles increases, the nodes exceed easily the range of transmission lead to link fault and the E2E Delay is getting higher, because we take more factors into consideration for calculating in order to select a better candidate.

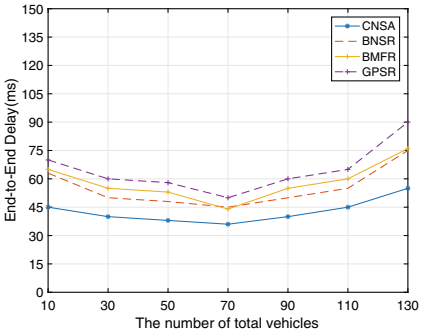


Fig. 5. The relationship of between E2E Delay and no. of vehicles

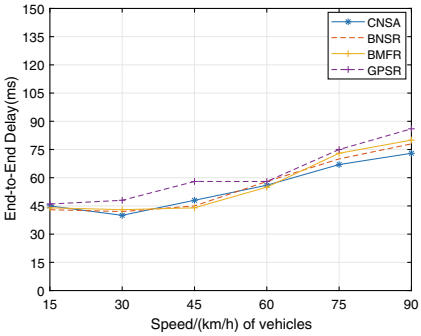


Fig. 6. The relationship of between E2E Delay and speed of vehicles

Packet Delivery and Performance Analysis. Figure 7 represents the relationship of number of nodes and the packet delivery rate. In the simulation environment of TABLE II. We can easily see the PDR are getting higher of all protocols when the number of vehicles less than 70. When the number of nodes increases to a certain value, there is a large variation in the concentration of adjacent section nodes. Then the retransmission packets increase which will make the packet delivery ratio has a downward trend. CNSA considers the stability of the link more, so the packet delivery ratio changes smoothly.

Figure 8 shows the effects of speed on packet delivery ratio. In the simulation environment of TABLE II. Because the network topology changes frequently, the data connection fractures easily. As the speed of vehicles increase, data packets in the network increase with the result of that a decline in packet delivery rate. CNSA has already considered the relative speed and ρ_{c_i} for neighbor nodes. According to the minimum W_{c_i} to choose the high stability of routing nodes which guarantees the reliable data transmission, so the performance of the packet delivery ratio is better than other protocols.

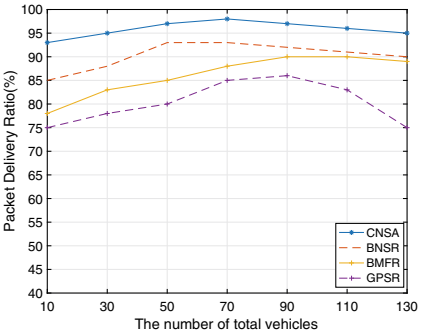


Fig. 7. The relationship of between PDR and no. of vehicles

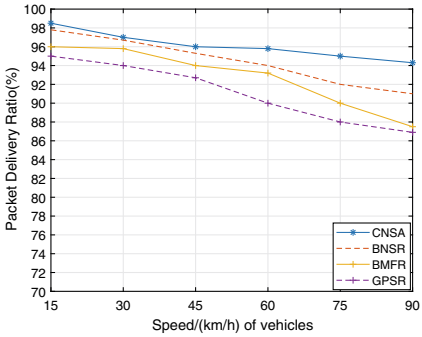


Fig. 8. The relationship of between PDR and speed of vehicles

5 Conclusion

In this paper, we present the algorithm for selecting the best available candidate node for VANETs by considering many aspects. Comparing with BNSR, BMFR, and GPSR, the result of simulation demonstrates that our routing protocol owns better packet delivery ratio and shorter end-to-end delay in the process of spreading packets from source node and destination node.

References

1. Al-Sultan, S., M.Al-Doori, M., Al-Bayatti, A.H., Zedan, H.: A comprehensive survey on vehicular Ad Hoc network. J. Netw. Comput. Appl. Elsevier, **37**, 380–392, January 2013
2. Karp, B., Kung, H.T.: GPSR: Greedy perimeter stateless routing for wireless networks. In: Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom), Boston, MA, August 2000, pp. 243–254 (2000)
3. IEEE 802.11 P Wireless Access for Vehicular Environment, Draft-Standard (2011). http://grouper.ieee.org/groups/S02/11
4. Wu, X., Tong, H., Mitton, N., Zheng, J.: MEDAL: a moving direction and destination location based routing algorithm for vehicular Ad Hoc networks. In: IEEE International Conference on Communications (ICC), Wireless Networking Symposium, pp. 6412–6416, June 2013, Budapest, Hungary
5. Mehan, B., et al.: BNSR: border node preferred social ranking based routing protocol for VANETs. In: Eighth International Conference on Contemporary Computing IEEE, pp. 555-560 (2015)
6. Raw, R.S., Lobiyal, D.K.: B-MFR routing protocol for vehicular ad hoc networks. In: International Conference on Networking and Information Technology (ICNIT), IEEE, pp. 420–430 (2010)

AP Placement in Indoor 3D TV White Space Network Based on Path Loss Model

Guanglin Zhang[1,2(✉)] and Zhengjian Yang[1,2]

[1] College of Information Science and Technology, Donghua University, Shanghai 201620, China
[2] Engineering Research Center of Digitized Textile and Apparel Technology, Ministry of Education, Shanghai 201620, China
glzhang@dhu.edu.cn , yangzhengjian@mail.dhu.edu.cn

Abstract. The increasing demand of wireless communication would lead to the scarcity of spectrum resources, and many countries and institutions have begun to release rules to use of TV white space (TVWS). In this paper, we presented an indoor three-dimensional TV white space network design. This is the first white space network design built in indoor 3D space. Considered the indoor environment seriously influences the wireless signal strength, we applied the path loss model. We proposed the 3D distance system and determined the AP placement based on path loss model. We built an experiment platform to communicate in TVWS, and we optimized AP placement in simulation. Experimental results show that this is an effectual architecture and might work in the real world. Our system satisfies the demand traffic rate and improves coverage and system throughput as a supplement.

Keywords: TV white space · Indoor 3D network · AP placement · Wireless communication · Path loss

1 Introduction

Nowadays, the raising popularity of wireless devices has led to an increasing demand for spectrum resources. According to the trend, the existing spectrums sooner or later will not be able to satisfy people's demand. Therefore, it is a direction to develop new spectrum resources for wireless communication in the future. In this context, dynamic spectrum access (DSA) presents a potential technique to unauthorized use the vacant spectrum to communicate without causing interference of licensed persons, so as to increase the utilization of spectrum resources and to cater to the growing demand of spectrum [1]. TV white spaces (TVWS), the unused portions of spectrum in the TV bands, refer to frequencies allocated to a broadcasting service but not used locally. The propagation characteristics of TVWS make it very suitable for wireless communication between devices. Because its spectrum is fairly low, the interference TVWS get in the process of transmission is much lower than other high-frequency spectra [2]. So TVWS

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 507–514, 2019.
https://doi.org/10.1007/978-981-13-6264-4_61

may spread farther and provide better service than other spectra in wireless communication.

Recently, the Federal Communications Commission (FCC) issued a final decree to allow unlicensed devices to use TV white space (TVWS) with no interfering to the licensed devices [3]. More and more study focused on TVWS, researchers began to focus on the practical application of TV white spaces [4,5]. Considering the requirement of indoor network communication is huge. Some researchers created a multi APs indoor white space network, WINET [6]. In view of the complexity of the indoor situation, such as the presence of obstacles, the attenuation of signal intensity caused by the wall and signal path loss when signals propagate from the transmitter to the receiver and other factors, we propose the indoor 3D TVWS network, the first design of TV white space based on path loss model.

In this study, we applied the path loss model to determine the coverage of signals, we selected the location of the node reasonably, and we designed the network architecture and optimized it. We first proposed a system scheme of 3D indoor TVWS network. Then, we studied and analyzed the methods of how to construct and build each part of the system. Finally, we optimize the AP layout and evaluate the system performance in the simulation.

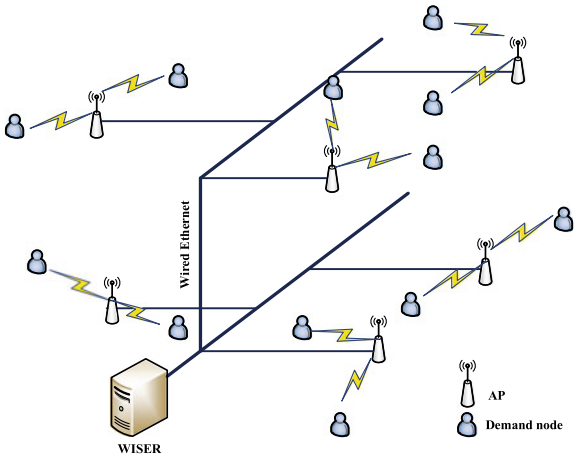


Fig. 1. Network architecture

2 Network Architecture

Indoor 3D white space network consists of three parts: APs, demand nodes, and WISER-indoor white space identification system [7]. WISER is an efficient white space recognition system, which is specialized for indoor conditions. We can obtain the indoor white space in different positions by querying it. The network architecture is shown in Fig. 1. Several APs are deployed in this network system and these APs are located on different floors and heights. Multiple APs cover

all demand nodes. All APs are connected to the wired ethernet network and WISER. One AP is able to associate multiple demand nodes.

3 System Model

We mainly focus on three-dimensional indoor architecture. We take path loss model into account when designing this network to make the system available and maximize performance as best as possible. In this work, we use WISER to draw a white space map and upload it to the database. Based on the white space map we obtained and the location of demand nodes, we get the potential AP placement points (AP candidate nodes). In order to achieve the best benefits, we take the maximum total throughput as the target and determine the AP placement.

3.1 Path Loss Model

Signals are emitted from the AP antenna so they emit a signal spreads in all directions in the building. A device point and a demand point can be represented as $p_i = (x_i, y_i, z_i)$ and $p_j = (x_j, y_j, z_j)$ in three-dimensional Euclidean space. Thus, the 3D distance between p_i and p_j is defined as

$$d(p_i, p_j)^{3d} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \tag{1}$$

Wireless signal strength is deeply affected by the environment, such as the wall materials, electronic devices, and obstructions. The log-normal shadowing model to describe the attenuation of the wireless signal strength is formally expressed as [8]

$$SS_{ijk} = SS_{0jk} - 10\beta \log(d_{ij}) - X_\sigma \tag{2}$$

where SS_{ijk} is the received power in dBm at demand i from the transmitter at node j on channel k, and SS_{0jk} is the original signal strength of AP node j on channel k. β is the path loss exponent; X_σ is a Gaussian random variable and standard deviation of σ in dB. The parameters β represent the effects of distance attenuation, and σ reflects environmental conditions. In our network design, AP nodes would satisfy the following condition:

$$SS_{ijk} \geq \gamma \tag{3}$$

where γ is the minimum received signal strength in dBm to meet the demand of wireless communication.

3.2 Determine AP Candidate Nodes

The AP nodes affect the use of the spectrum and the performance of system. Moreover, the number of AP should be reasonable because too much AP can cause excessive costs and can cause serious interference with each other. We

try to ensure that all demand nodes are covered with as possible as less APs. In our design, each AP covers at least two demand nodes. The signal strength received at each demand point is greater than the minimum received power γ for adequate wireless communication. As shown in Fig. 2, we choose positions to meet the above conditions as AP candidate nodes.

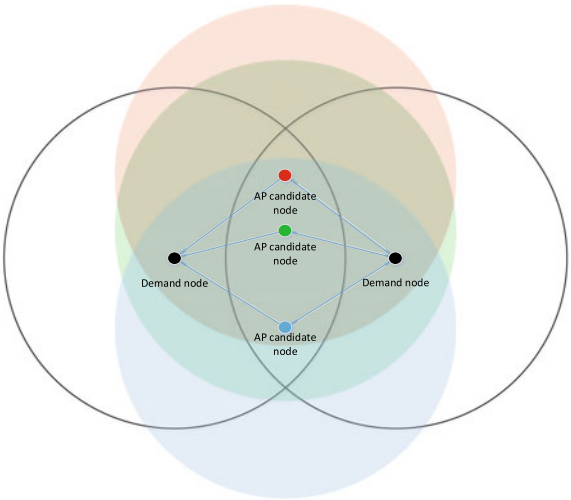


Fig. 2. Determine AP candidate node

Algorithm 1 Determining AP candidate nodes

calculate d_{ij} satisfy $SS_{ijk} = \gamma$
for all demand nodes i **do**
  Make spheres of radius d_{ij};
  Find all node n satisfy $M_{nk} = 1$ in overlap areas of all spheres;
end for
Find all AP candidate nodes $j \in n$;
Impose the constraint:
1.Channel k is common vacant at both AP candidate node j and demand node i;
2.$SS_{ijk} \geq \gamma$

3.3 Model Formulation

In order to build a three-dimensional indoor white space network, we get the following information. The signification of notations is shown in Table 1.

To indicate the installation of the AP, we define variable X_{ijk} to represent the relationship among AP candidate nodes, demand nodes, and TV channels. Express as follows:

$$X_{ijk} = \begin{cases} 1, & \text{if } i \text{ is covered by AP at } j \text{ on channel } k \\ 0, & \text{otherwise} \end{cases}$$

Table 1. Main notations

Notation	Definition
i	Demand node
j	AP candidate node
k	TV channel
a_i	Set of demand node
a_j	Set of AP candidate node
d_i	Average traffic demand of node i
ρ	The maximal traffic rate of single AP
N	Number of available APs
X_{ijk}	Association decision variable
T_{ijk}	Estimated actual throughput

We have selected the location of AP candidate nodes and confirmed the number of offered APs. The next step is to determine the specific AP placement. We use T_{ijk} to represent the estimated actual throughput of each link. We define the AP placement problem as a **M**aximizing **S**ystem total **T**hroughput problem (**MST**). The MST problem is formulated as follows:

$$\text{Max} \quad \sum_i \sum_j \sum_k T_{ijk} X_{ijk} \tag{4}$$

s.t.

$$X_{ijk} \leq N, \forall i \in a_i \tag{5}$$

$$X_{ijk} = 1_{\{T_{ijk}>0\}} \tag{6}$$

$$\sum_j \max_k X_{ijk} = 1, \forall i \in a_i \tag{7}$$

$$\sum_i \sum_j \sum_k X_{ijk} \cdot T_{ijk} \geq a_i \cdot d_i \tag{8}$$

$$\sum_i \sum_k X_{ijk} \cdot T_{ijk} \leq \rho, \forall j \in a_j \tag{9}$$

The objective function (4) is to maximize the total throughput of the network system by a reasonable AP placement. Constraint (5) indicates that the number of APs to be placed must be no greater than the number of offered APs. Constraint (6) shows the mapping relationship between X_{ijk} and T_{ijk}, because T_{ijk} describes the existence of the links. If $T_{ijk} > 0$, then the link must exist ($X_{ijk} = 1$). If $T_{ijk} = 0$, the link not exists ($X_{ijk} = 0$). Formula (7) specifies each demand node can establish a wireless connection with only one AP. Constraint (8) guarantees that the total actual throughput can satisfy the total throughput demand in the system. Constraint (9) describes the total throughput of the

demand notes connected with the same one AP cannot exceed the maximum downlink traffic rate of this AP. We realized our system through simulating in later sections.

4 Performance Evaluation

4.1 Communication with USRP on TVWS

In order to study TVWS, we set up a platform for experiment.

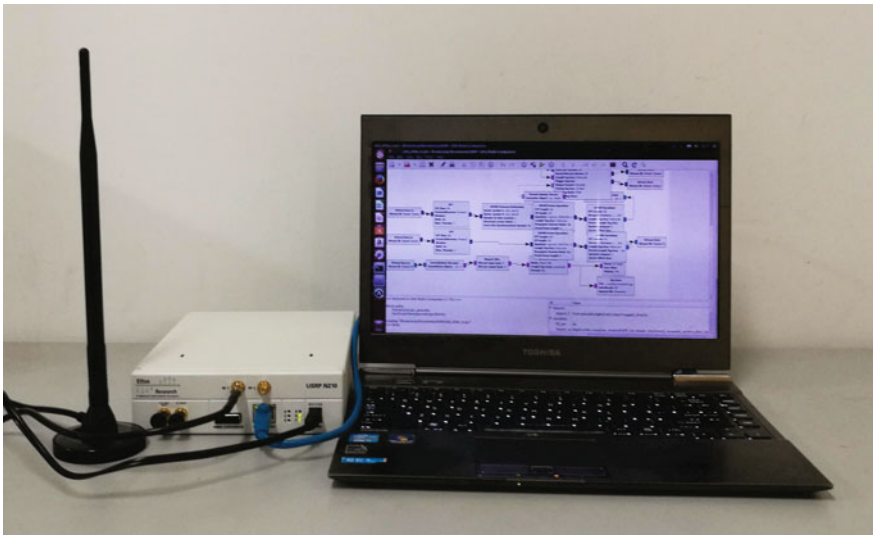


Fig. 3. Devices of platform

As shown in Fig. 3, laptops and universal software radio peripheral (USRP) devices are used to measure the spectrum and deliver a picture on TV white space. The software environment we adopted is GNU Radio and UHD. We used WBX 50-2200Mhz Rx/Tx daughter board. We tried to use TVWS for communication. In our program, each USRP is connected to one laptop. One is set for transmitting data and the other is for receiving it. We used OFDM modulation to transmit signals. The center frequency we set is 570Mhz, which is unoccupied in our measurement. We transmitted a picture with a USRP and received with the other one. The received signal strength is -60 dB; we can see that the received signal strength is similar to Wi-Fi. The experiment shows that TV white space is entirely feasible for wireless communication.

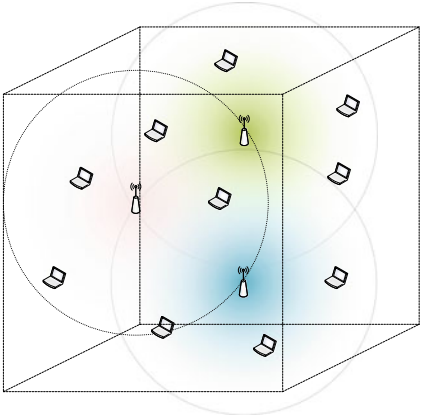


Fig. 4. The schematic diagram of system simulation

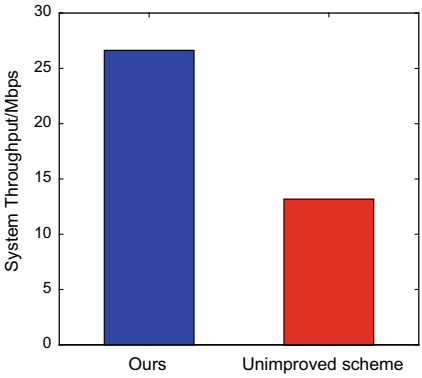


Fig. 5. Our scheme improves system throughput

4.2 Network Simulation

Our main objective is to maximize the system throughput by optimizing the AP placement. In order to carry out our research, we set up an experimental platform on Network Simulator version 2, and a multi-node network is simulated. In our experiment, 10 demand nodes need to be covered. As shown in Fig. 4, we set the bandwidth for each AP to 8Mbps and the carrier frequency within TV white space. Several wireless connections are established between APs and demand nodes, and UDP protocol is used to transfer data. In this process, we measured and analyzed the relevant parameters. In terms of coverage and throughput, we compared with unimproved scheme. In Fig. 5, the result shows that we can improve system total throughput exponentially.

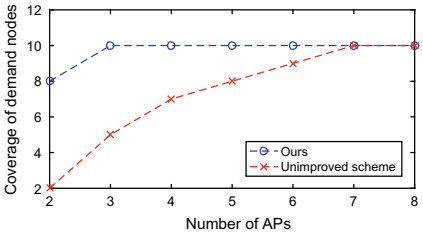


Fig. 6. Relationship between coverage and number of APs

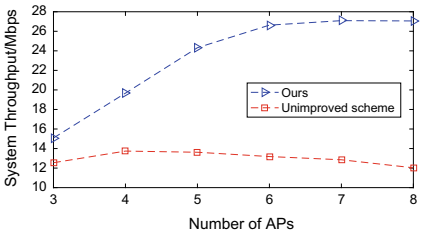


Fig. 7. Throughput changes with the number of APs

The mapping between coverage and AP number is shown in Fig. 6. We can cover all demand nodes with only 3 APs while unimproved scheme needs at least 7 APs. And Fig. 7 displayed that whether in our design or in other scenarios, as

the number of APs continues to increase, system throughput will grow slowly or even decline. This is due to the increasing interference among APs. So it is necessary to choose the most reasonable AP number. Otherwise, spectrum allocation is also considered in our research. We find that if the spectrum is allocated in a separated way, the total throughput of the system would be greater than the other case. Contrarily, if the same or similar channels are distributed to nearby APs, the system total throughput might drop to a very low value. So in order to realize the best performance and the fairness among users, the channel of wireless links we set up for each AP should be as far apart as possible.

5 Conclusion

We propose a new TV white space network design built in 3D indoor environments in this paper. We applied the path loss model in wireless communications. During the experiment, we realized communication in TVWS with USRP devices. The optimization problem of AP placement is solved. The result shows the system we design satisfies the requirement of a wireless network in both coverage and system total throughput, and it is an effective and practical system. There are still many directions to be researched and explored in the future, for example, how to dynamically allocate the spectrum to users according to actual situation. It is very significant to set up an indoor TV white space network in the real world.

References

1. Hossain, E., Niyato, D., Han, Z.: Dynamic Spectrum Access and Management in Cognitive Radio Networks. Cambridge University Press (2009)
2. Flores, A.B., Guerra, R.E., Knightly, E.W., Ecclesine, P., Pandey, S.: Ieee 802.11 af: A standard for tv white space spectrum sharing. IEEE Commun. Mag. **51**(10), 92–100 (2013)
3. Magazine, R.: Fcc adopts rules for unlicensed use of television white spaces. Retrieved at p. 4 (2008)
4. Borth, D., Ekl, R., Oberlies, B., Overby, S.: Considerations for successful cognitive radio systems in us tv white space. In: 3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2008. DySPAN 2008, pp. 1–5. IEEE (2008)
5. Bahl, P., Chandra, R., Moscibroda, T., Murty, R., Welsh, M.: White space networking with wi-fi like connectivity. ACM SIGCOMM Comput. Commun. Rev. **39**(4), 27–38 (2009)
6. Zhang, J., Zhang, W., Chen, M., Wang, Z.: Winet: indoor white space network design. In: 2015 IEEE Conference on Computer Communications, pp. 630–638. IEEE (2015)
7. Ying, X., Zhang, J., Yan, L., Zhang, G., Chen, M., Chandra, R.: Exploring indoor white spaces in metropolises. In: Proceedings of the 19th annual International Conference on Mobile Computing & Networking, pp. 255–266. ACM (2013)
8. Rappaport, T.S., et al.: Wireless Communications: Principles and Practice, vol. 2. Prentice Hall PTR New Jersey (1996)

Modified Coordinated Multipoint Transmission in Ultradense Networks

Jiajun Zhang(✉), Xuanli Wu, Wei Wu, and Yutong Xiao

Communication Research Center Harbin Institute of Technology, Harbin, China
{zhangjiajun_hit,hitadao}@163.com,{xlwu2002,kevinking}@hit.edu.cn

Abstract. Ultradense network (UDN) is considered as one of the key technologies in the future 5G network. The system capacity could be increased greatly with the cell densification. It can be realized by the deployment of small cells which have low transmit power and coverage. However, the densification of small cells can also bring strong inter-cell interference (ICI) which limits the system performance. Coordinated multipoint transmission (CoMP) can eliminate interference or transform the interference into useful signals. It is introduced to improve the cell edge users' performance in the traditional sparse network, while it has some constraints in the dense network. In this paper, we propose a modified CoMP which can improve the performance of ultradense network. Users choose multiple small cells according to two different thresholds, and the small cells serve the users by transmitting signals or just mute. The simulation results show that this method can improve the system energy efficiency with little data rate reduction.

Keywords: Ultradense Networks · CoMP · Interference coordination

1 Introduction

With the continuous increasing of the number of smart devices and user needs, the traffic of the wireless communications is increasing rapidly. In order to face the challenges, the next-generation wireless communication system (5G) will be deployed around 2020. Compared with current wireless communication system (4G), 5G has higher requirements, and one of the most urgent requirements is system throughput improvement [1]. In order to achieve this goal, a variety of technologies have been proposed. All methods are performed from one or more of the three aspects: providing more spectrum resources, increasing the spectrum efficiency of each cell, or increasing the cell density.

Ultradense network has been concerned as one of the key technologies in the next-generation network [2]. Since the beginning of the mobile communications industry, cell fragmentation and densification has been one of the most effective ways to increase system capacity and user demands. In the past few decades, the increase in spectrum efficiency due to the reduction of the area of the cell and the increase in the spatial multiplexing rate have been enormous. Different from the

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 515–521, 2019.
https://doi.org/10.1007/978-981-13-6264-4_62

traditional spare network, a large number of small cells (SC) are deployed at the hot spot areas to share loads of the macro cell or at the blind areas to enhance cell coverage. The small cells are cheap, easy to deploy with low transmit power, and small coverage areas which means the pass-loss is small.

However, the densification of the network would cause severe inter-cell interference. To deal with the ICI and improve the cell edge users' performance, coordinated multipoint transmission is introduced in the traditional spare network. In UDN the interference is more complicated. User-centric interference coordination is considered a feasible solution [3]. Similar to the CoMP, the virtual cell concept is introduced [4]. But they just consider the best cells as the coordination cells and the severing cells are all treated equally. The interference coordination methods would not fully eliminate the interference or increase the system complexity along with the energy consumption.

In this paper, we introduced a modified coordinated multipoint transmission method that could take advantage of the coordination and reduce the complexity. The rest of the paper is arranged as follows: the second section presents the system model, the third section introduces the modified CoMP model in UDN, the fourth section introduces the simulation parameters and the simulation results are analyzed, and the fifth section concludes this paper.

2 System Model

In ultradense networks, the macro cell covers the whole region while the small cells are distributed at the macro cell blind areas and the hot spot areas. For research, the macro cell is located at the center of the system and the small cells are distributed follows the Poisson point process (PPP) [5]. To avoid the cross-tier interference, the macro and small cells are using different frequency resources. Considering the downlink OFDMA system, all the users in the same cell are using orthogonal resource to avoid intra-cell interference. All the small cells are equipped with single omit antennas with the same maximum transmission power and reuse the same resource. There are totally L subcarriers, which means there are L resource blocks (RB). For the channel model, the pass-loss factor is 4 and the small-scale fading follows the Rayleigh's fading.

Users select their own serving nodes according to the average signal power received over a period of time, thus obtaining the maximum signal-to-noise ratio (SNR). The judgment is simple and widely used which is generally called reference signal received power (RSRP). Each user would select at most M small cell as their serving cells according to the RSRP.

Let B_Ω and U_Ω be the small cell set and user set, for any arbitrary user $i \in U_\Omega$, the severing cell(s) could be the macro cell or small cell(s). Let $c_{b,i} \in \{0, 1\}$ represent whether cell $b \in B_\Omega$ is the severing cell of user i. If cell b is the severing cell of user i, then $c_{b,i} = 1$, otherwise $c_{b,i} = 0$. $\sum_{b \in U_\Omega} c_{b,i} = 0$ means that the user i is served by the macro cell.

If $\sum_{b \in U_\Omega} c_{b,i} \neq 0$ which means user i is severed by small cell(s), the SINR of user i at a specific RB is

$$SINR_i = \frac{\sum_{b \in B_\Omega} c_{b,i} P_{b,i} |h_{b,i}|^2}{\sum_{b \in B_\Omega} (1 - c_{b,i}) P_{b,i} |h_{b,i}|^2 + \sigma_z^2} \tag{1}$$

where $h_{b,i}$ is the channel coefficient of cell b and user i, $P_{b,i}$ is the transmission power of cell b, and σ_z^2 is the power of additive white Gaussian noise. So the data rate of user i at the RB is

$$C_i = B * \log_2 (1 + SINR_i) \tag{2}$$

To simplify the system, we assume that all the UEs severed by the macro cell have the same data rate R_0. Then the sum of the data rate of the system is

$$C = \sum_{i \in U_\Omega} C_i + n_M * R_0 \tag{3}$$

where n_M is the number of UEs severed by the macro cell. The total power consumption P_c can be calculated as follows:

$$P_c = \sum_{b \in B_\Omega} c_{b,i} P_{b,i} + P_M \tag{4}$$

where P_M is the power consumption of the macro cell. The energy efficiency can be defined as

$$EE = \frac{C}{P_C} \tag{5}$$

3 Modified CoMP in UDN

In traditional CoMP, the user equipments (UEs) choose multiple base stations (BSs) as the severing BSs. The severing BSs transmit the same signals via the same resource or there is just one BS transmits the signals while other BSs keep muting on that resource. CoMP is introduced to improve the cell edge users performance. In the traditional spare networks, the cell edge users receive low signal power but strong interference. The interference is mainly from the neighbor cell, which means there exists a dominated interference at the cell edge. The dominated interference ratio(DIR) is introduced:

$$DIR = \frac{I_{strongest}}{I_{strongest} + I_{others}} \tag{6}$$

In traditional spare networks, the DIR is high, and the interference coordination is mainly focused on dealing with the dominated interference. However in

UDN, the DIR may be so small which means that there are several interference sources and the difference of interference power is small. The coordination points are small which is barely large than 3. So the method to eliminate the dominated interference would not work well in UDN [6].

With CoMP, the users select severing small cells according to an RSRP threshold, if the threshold is high some of the interference cells may not be selected, while if it is low there may be too many cells selected, which would lead to a high cost but low gain. Different from the joint transmission (JT) CoMP that all the coordinated cells transmit signals to the UEs and the dynamic cell selection (DCS) CoMP that just one main severing cell transmit signals while other cells keep muting, we take advantage of both and propose a modified CoMP method in UDN. By which the UEs choose several BSs as the severing base stations according to the average RSRP. There are two different thresholds $P_{severing}$ and $P_{connecting}$ with the constraints:

$$P_{connecting} \leq P_{severing} \tag{7}$$

If the RSRP of a small cell b is larger than $P_{severing}$, $c_{b,i} = 1$ and the cell b transmits signals to the UE i. If the RSRP of a small cell b is larger than $P_{connecting}$ but small than $P_{severing}$, $c_{b,i} = 1$ and the cell b connects with the UE i but it would not transmit signals and just keeps muting on the specific RB which means $P_{b,i} = 0$. If the RSRP is smaller than the $P_{connecting}$, $c_{b,i} = 0$ and the UE i is not in the coverage of that small cell b.

The muting small cells have contributed little to the total signal power, if they transmit signals to the UEs the data rate would not increase so much and the power would be wasted the system complexity would be high. However, if these cells severe other UEs using the same resource, it would cause interference that is high enough to decrease the data rate severely. In order to balance the data rate and the power assumption, these small cells would be one of the severing cells of the UEs, but there would be no signal transmitted.

4 Simulation Results

We use the property fairness (PF) algorithm as resource allocate algorithm, and the water filling algorithm as the power allocate algorithm. The simulation parameters are shown in Table 1.

The system sum data rate is shown as Fig. 1. The full transmission theme that all the connected cells transmit signals to the UE and the connection threshold is low enough to eliminate the interference. The no muting theme considers a higher threshold to save energy but the interference level would be high. It is shown that the modified CoMP and full transmission theme can improve the data rate a lot compared with the no muting theme. With the modified CoMP theme, the data rate decreases a little compared to the full transmission theme.

Table 1. Simulation parameters

Name	Symbol (unit)	Value
Frequency	f_c (GHz)	3
Band	B(MHz)	20
RB number	L	25
Small cell power	P (dBm)	10
Small cell coverage	(m)	30
Small cell density	λ_s (km^2)	100∼1000
User density	λ_u (km^2)	800
Area	S (km^2)	1
Maximum cells UEs can connect	M	4

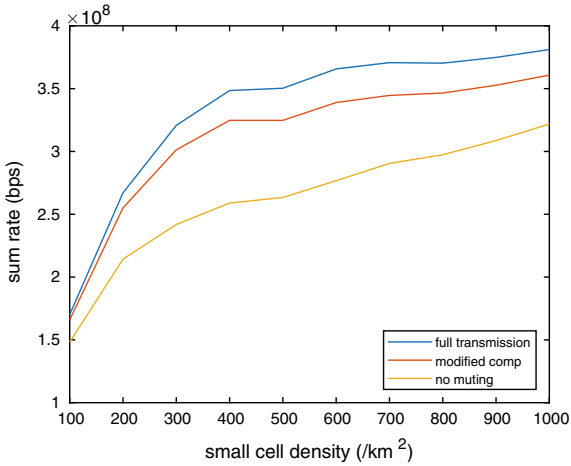


Fig. 1. System data rate

The rate increases as the small cell density increases when the density is small, this is because the more small cells there are, the more severing cells a UE can have, and the signal power gets higher further improves the data rate.

However, it can be found in Fig. 2 that compared with the full transmission theme, the power consumption of modified CoMP theme reduces a lot. That is, though the modified CoMP decreases the data rate a little, it can decrease the power consumption a lot, which leads to a higher energy efficiency. The energy efficiency curves are shown in Fig. 3.

The system energy communication increases and the energy efficiency decreases as the small cell density increases. That is, from the perspective of energy efficiency, the network density should not be too high.

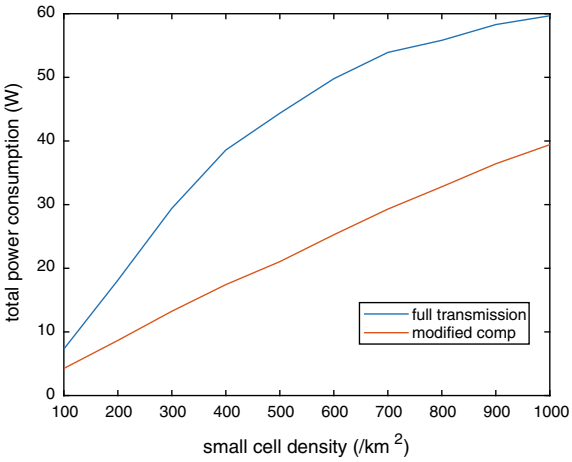


Fig. 2. System power consumption

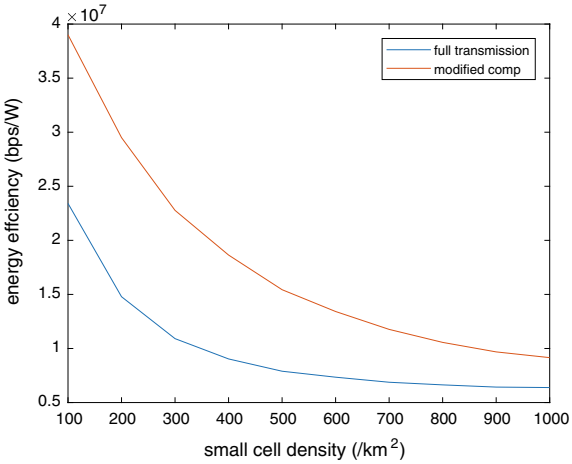


Fig. 3. System energy efficiency

5 Conclusion

In this paper, we propose a modified coordinated multipoint transmission method in UDN. The users choose the severing small cells according to two different thresholds. Some of the small cells transmit signals and some just mute. The muting small cells have contributed a little to the total signal power but can reduce the interference power a lot. Comparing to the unmuting theme, the system data rate increases, and compared with the full transmit theme, the system data rate gets little decreased but the energy assumption gets decreased, and a higher energy efficiency could be obtained.

Acknowledgements. This research is supported by Natural Science Foundation of China (Grant No.61501136).

References

1. Xiaohu, G., Song, T., Guoqiang, M., Chengxiang, W., Tao, H.: 5G ultra-dense cellular networks. IEEE Wirel. Commun. **23**(1), 72–79 (2016)
2. Kamel, M., Hamouda, W., Youssef, A.: Ultra-dense networks: a survey. IEEE Commun. Surv. Tutor. **18**(4), 2522–2545 (2016)
3. User-Centric Intercell Interference Coordination in Small Cell Networks: IEEE International Conference on Communications, pp. 5747–5752 (2014)
4. Chiyang X., Jie, et al.: Downlink transmission scheme based on virtual cell merging in ultra dense networks. In: Vehicular Technology Conference (VTC-Fall) (2016)
5. Andrews, J.G., Baccelli, F., Ganti, R.K.: A tractable approach to coverage and rate in cellular networks. IEEE Trans. Commun. **59**(11), 3122–3134 (2011)
6. Xiaoyi, W., Eugene V., Amitava, Ghosh.: Dynamic cell muting for ultra dense indoor small cell deployment scenario. In: IEEE International Conference on Communication Workshop, pp. 148–153 (2015)

Research on Strategy of Keeping Multi-UAVs Flight Formation Based on Distribution Method

Mao Yi[1], Tong Ming[2], and Liu Rong[1(✉)]

[1] UAV Research Institute, Nanjing University of Aeronautics and Astronautics, Nanjing, China
liurong@nuaa.edu.cn
[2] State Key Laboratory of Air Traffic Management System and Technology, Nanjing, China

Abstract. Concerning the strategy of keeping multi-UAVs flight formation based on distribution method, first, large-scale multi-UAVs formation is divided into numerous two-UAV sub-formations in the way of "hierarchy". Then, lead and wing aircraft control approach is adopted in sub-formations, i.e., UAVs in the formation are classified into lead and wing aircrafts, which are complementary to each other. In this way, formation control is thus realized by controlling wing aircrafts that track the lead aircraft. The close connection of sub-formations also constitutes a tracking chain–UAV i tracks UAV i-1. Finally, multi-UAVs formation flight is realized.

1 Introduction

Control system design for multi-UAVs (unmanned aerial vehicle) formation flight is different from that for single-UAV [1]. The stability of a single aircraft is controlled, so is the position and distance relative to the lead aircraft or a selected reference point in formation. Therefore, in mission fulfillment by making use of UAV flight, formation design and selection is of great importance [2–4]. In addition to impacts of aerodynamic force between UAVs, mission demands and redundancy of internode information exchange should also be taken into consideration. Formation collision in maneuver should also be avoided.

UAV flight formation based on distribution method is shown in Fig. 1. UAV1# is lead aircraft. UAV2# and UAV4# follow UAV1# in flight, with their relative position unchanged so as to ensure their stable position in the formation. UAV3# can realize its position unchanged in the whole formation only by mastering details of UAV2# and keeping its relative position stable [5–7].

Despite several advantages of distribution approach, relative offsets caused by dynamic effect may differ if there are multiple UAVs surrounding the aircraft in the formation. As a result, non-matching and inconsistency of each UAV's expected

Research supported by AB Foundation.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 522–529, 2019.
https://doi.org/10.1007/978-981-13-6264-4_63

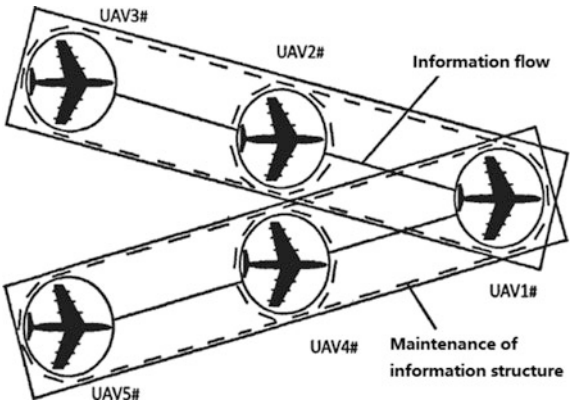


Fig. 1. Distributed formation keeping strategy

correction may appear. How to extract the right correction for formation keeping from inconsistent corrections and figure out the precise state must be solved [8].

2 Lead and Wing Aircraft Following Modeling

To summarize the motion principle of UAV formation based on distribution method, wing aircraft receives lead aircraft's information of velocity, direction, and height via data chain first [9]. Then, according to relative motion equation and formation order, based on its flight state, wing aircraft works out its own velocity, direction, and height with formation controller so as to realize flight in formation. Therefore, it is necessary to primarily investigate into multi-UAVs formation control modeling [10].

(1) Lead aircraft's mathematical model

Lead aircraft's mathematical model is set up in an inertial coordinate system, as shown in Fig. 2. It is a fixed reference coordinate system that is used to describe lead aircraft's motion track.

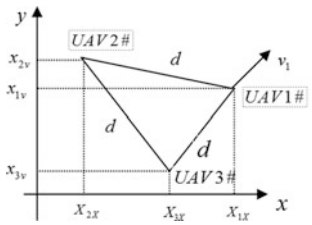


Fig. 2. Inertial coordinate system

In Fig. 2, UAV1# is lead aircraft. UAV2# and UAV3# are, respectively, left and right wing aircraft. $x_{ix}(i = 1, 2, 3)$ is UAVi's position along the x-axis; and $x_{iy}(i = 1, 2, 3)$ is UAVi's position along the y-axis. Lead aircraft's (UAV1#) mathematical model is

$$\begin{cases} \dot{x}_{1x}(t) = v_{1x}(t), \dot{v}_{1x}(t) = f_x(t) + u_{1x}(t) \\ \dot{x}_{1y}(t) = v_{1y}(t), \dot{v}_{1y}(t) = f_y(t) + u_{1y}(t) \end{cases}$$

Thereinto, $v_{1x}(t)$ and $v_{1y}(t)$ are the projected velocity of lead aircraft along the x- and y-axes. $f_x(t)$ and $f_y(t)$ are unknown disturbances. $u_{1x}(t)$ and $u_{1y}(t)$ are control input of lead aircraft along the x- and y-axes.

(2) Wing aircraft's mathematical model

Wing aircraft's mathematical model is set up in a velocity coordinate system, as shown in Fig. 3. It is a reference coordinate system in motion, in which lead aircraft's position and velocity are, respectively, defined as the system's original point and vertical axes.

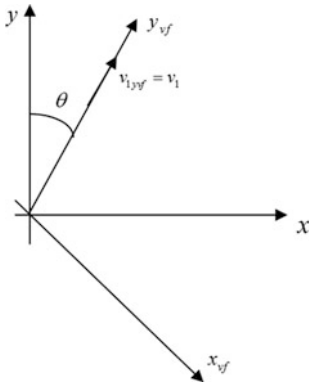


Fig. 3. Velocity coordinate system

Wing aircraft's (UAV1#, i = 2, 3) mathematical model is

$$\begin{cases} \dot{x}_{ixvf}(t) = v_{ixvf}(t), v_{ixvf}(t) = f_{ixvf}(t) + u_{ixvf}(t) \\ \dot{x}_{iyvf}(t) = v_{iyvf}(t), v_{iyvf}(t) = f_{iyvf}(t) + u_{iyvf}(t) \end{cases}$$

3 Solution of Inconsistent Expected Correction

UAVs may deviate from the expected position in flight. According to the number of deviated UAVs and offsets, the formation can be classified into four states. For each state, the requirement differs.

(1) In sound state. No adjustment is needed.
(2) Several UAVs deviate. Disturbance attenuation in a small scope is expected. The non-deviated UAVs avoid from impacts as much as possible.
(3) A large number of UAVs deviate in a small scope. Each UAV is expected to be stable as much as possible when returning to the original position. No disorder is required.
(4) Messy formation. In this state, quick formation recovery is required. UAVs apply partial formation state, take corresponding actions, and keep formation via inter-UAV coordination.

UAV formation composed of N UAVs is shown in Fig. 4. Surrounding UAV $m(m = 1, 2, 3 \ldots N; m \leq N)$ is UAV $Km(Km \leq N - 1)$. Vector of relative position offset between UAV $i(i = 1, 2, 3 \ldots Km)$ and UAV m, $\vec{C}_i^m = (\Delta X_i^m, \Delta Y_i^m, \Delta Z_i^m)$ (ΔX_i^m, ΔY_i^m, ΔZ_i^m are, respectively, three components of $\vec{C}_i^m$ in the coordinate of flight path.

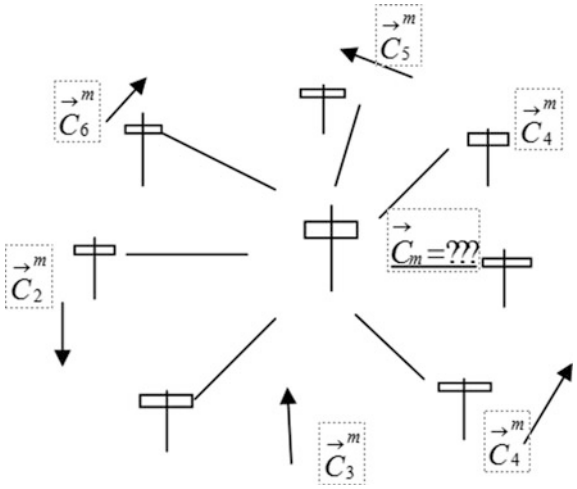


Fig. 4. Formation of UAVs

Therefore, the following requirements on formation keeping strategy are raised:

(1) UAV motion is decided based on partial formation state;
(2) Extract $\vec{C}_m$ by using $\vec{C}_L^m$ and inconsistent $\vec{C}_i^m$;
(3) Judge the formation state and take corresponding actions;
(4) Briefly deal with dynamic performances in flight.

In vector of relative position offsets $\vec{C}_i^m$, some are advantageous for keeping formation, while others may bring disadvantageous factors. Unselective use will cause disadvantageous $\vec{C}_i^m$. For example, commonly adopted equalization approach can balance all $\vec{C}_i^m$ in the zone, but some UAVs will be affected by a minority of UAV's disturbance, and then in a larger scale.

For the purpose of finding out an appropriate $\vec{C}_i^m$, it is necessary to find out a reference point and judge value of each $\vec{C}_i^m$. The point with least disturbance impacts in the zone can be searched as the reference point, which, however, is complex in calculation and not suitable for formation keeping in UAV maneuver. Furthermore, judgment based on detailed information in partial zone cannot ensure correct judgment of overall state in the whole zone. Therefore, it is not necessary to find out the most stable reference point in the partial zone. In formation flight, there is little chance of same offset vector among different UAVs. Coupled with the simplicity in finding out $\min(\vec{C}_i^m)$ (minimal offset of relative position), it is a proper method to apply $\min(\vec{C}_i^m)$ as the basis of the value of $\vec{C}_i^m$.

4 Judgment of State of Formation Keeping

Due to aerodynamic impacts on UAVs in the formation, their dynamic effects are coupled. When the ratio of AR (formation distance) to b (wingspan) is larger than 3 (included), the formation is scattered, with weak coupling effect. If the offset is within the controller's error control scope ε, it can be regarded as in a sound state. When $AR/b < 3$, formation is characterized by intense coupling effect and close connection. In this state, one UAV's offset will affect control of other UAVs. The formula of estimating aerodynamic induction velocity of coplanar two-UAV flight deduced according to Biot–Savart theorem is as below:

$$q = \frac{\Gamma}{4\pi h}(1 - \cos\beta)$$

$$\Gamma = \frac{L}{\rho v_m b_v} = \frac{2 v_m b C_L}{\pi AR}$$

The ratio of induction velocity q and flight velocity v_∞ is directly proportional to b, but inversely proportional to square of AR. Suppose that the controller's tolerance limit for the ratio is δ, considering cumulated influences of multi-UAVs, set b's change within the scope of $\delta/2$ as tolerable offset. When the offset is smaller, then it can be considered in a sound state. Above all, the basis of sound state judgment is as follows: When $AR/b \geq 3$, $\vec{C}_L^m \leq \varepsilon AR, \max(\vec{C}_i^m) \leq \varepsilon AR$; when $AR/b < 3, \vec{C}_L^m \leq \delta * b/2$, $\max(\vec{C}_i^m) \leq \delta * b/2$. If the formation is not in sound state, further judgment is needed. In general, when $\vec{C}_i^m \geq AR/2$, UAV i has been beyond the formation of UAVm. $\vec{C}_i^m$ that is smaller than $AR/2$ is regarded as relative position offset inside the formation (Fig. 5).

If $AR/b \geq 3$ and $\min(\vec{C}_i^m) \leq \varepsilon AR$, or $AR/b < 3$ and $\min(\vec{C}_i^m) \leq \delta * b/2$, it implies that at least one UAV is in sound state in the formation. Considering that "there is little chance of same offset vector among different UAVs", it can be regarded that the UAV and that corresponding to "$\vec{C}_i^m \leq AR(\delta * b/2)$" has no offset. The formation is in the state of "only a minority of UAVs have offsets" under this condition. The UAV should continue keeping the formation without taking into consideration UAVs' with large offsets.

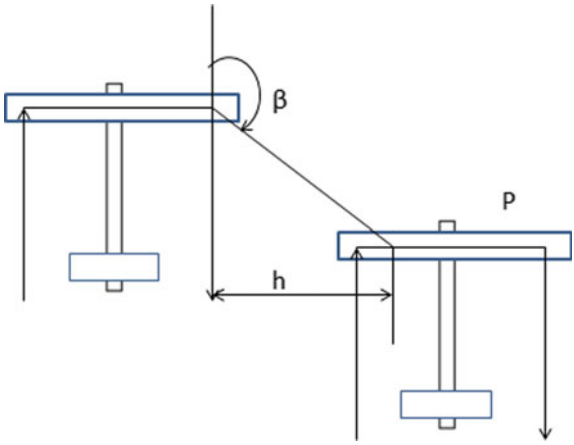


Fig. 5. Relativeness of coplanar two-UAV flight

If $AR/b \geq 3$ and $\varepsilon AR \leq \min(\vec{C}_i^m) \leq AR/2$, or $AR/b < 3$ and $\delta * b/\overrightarrow{2} \leq \min(\vec{C}_i^m) \leq AR/2$, and $\max(\vec{C}_i^m) < AR/2$, it indicates that each surrounding UAV has relative offset to different degrees within the scope of formation, which is in the state of "a number of UAVs have offsets in a small scope," without a reference point. Therefore, all $\vec{C}_i^m$ should be taken into consideration to ensure the stability of formation. Large offsets shall adopt a small ratio, while small offsets shall adopt a large one. If $\max(\vec{C}_i^m) \geq AR/2$, it means that some UAVs are beyond the formation, which is not advantageous for formation keeping. Therefore, the UAVs shall be eliminated in subsequent dealing.

If $\min(\vec{C}_i^m) \geq AR/2$, it means that surrounding UAVs have been beyond the formation and that the formation is scattered. Under this condition, the formation is required to be quickly rearranged. UAVs with the minimal relative position offset can be arranged with the UAV in a formation in the shortest time. Therefore, the UAVs corresponding to $\min(\vec{C}_i^m)$ should be considered so that the overall formation can realize quick and orderly rearrangement.

5 Distributed Keeping Strategy Algorithm

According to solution of inconsistent expected correction and model of formation keeping state judgment approach, the distributed keeping strategy algorithm is as below:

(1) Measure the vector of relative position offset between the UAV (UAV m in the formation) and lead aircraft and that between the UAV and the surrounding UAV, $\vec{C}_i^m$ and $\vec{C}_1^m, \vec{C}_2^m \ldots \vec{C}_{gm}^m$. $\vec{C}_i^m = (\Delta X_i^m, \Delta Y_i^m, \Delta Z_i^m)$, $\vec{C}_L^m = (\Delta X_L^m, \Delta Y_L^m, \Delta Z_L^m)$;

(2) Calculate $\min(\vec{C}_i^m)$ and $\max(\vec{C}_i^m)$;

(3) If $AR/b \geq 3 \cap \left[\vec{C}_L^m \leq \varepsilon AR \cup \max(\vec{C}_i^m) \leq \varepsilon AR\right]$ is true, or $AR/b < 3 \cap \left[\vec{C}_L^m \leq \delta * b/2 \cup \max(\vec{C}_i^m) \leq \delta * b/2\right]$ is true, formation correction, $\overrightarrow{C}_m = 0$. If they are all false, go to step (4);

(4) If $AR/b \geq 3$ and $\min(\vec{C}_i^m) \leq \varepsilon AR$ is true, or $AR/b < 3$ and $\min(\vec{C}_i^m) \leq \delta * b/2$ is true, correction $\vec{C}_m = [\min(\Delta X_i^m), \min(\Delta Y_i^m), \min(\Delta Z_i^m)]$. If they are all false, go to step (5);

(5) If $AR/b \geq 3$ and $\varepsilon AR \overrightarrow{\leq} \min(\vec{C}_i^m) \leq AR/2$ is true, or $AR/b < 3$ and $\delta * b/\overrightarrow{2} \leq \min(\vec{C}_i^m) \leq AR/2$ is true, $\vec{C}_m = \min(\overrightarrow{aC}_1^m + b\vec{C}_2^m + \ldots \overrightarrow{\tau C}_j^m)$. There into, $(\alpha + \beta + \cdots + \tau) = 1, (\vec{C}_1^m, \vec{C}_2^m, \ldots \vec{C}_j^m) < AR/2$,$(j = 1, 2, 3 \ldots Km, j \leq Km,$ when $\max(\vec{C}_i^m) < AR/2, j = Km)$. If they are false, go to step (6);

(6) If $\min(\vec{C}_i^m) \geq AR/2$ is true, $\vec{C}_m = \min(\vec{C}_i^m)$. If it is false, report the error and repeat step (1).

6 Conclusion

When multi-UAVs network-based collaboration is needed to complete online missions synchronously, distributed control structure can be applied, with such advantages as simplicity, reliability, a small amount of information, and easy avoidance from information conflicts. From the perspective of engineering, distributed structure can be more easily realized and maintained. In addition, with strong strategic adaptability, outstanding extensibility and fault tolerance of distributed control strategy, by interacting its position, velocity, posture and goal of motion with nearby UAVs in the formation, each UAV can coordinate with each other to complete multi-UAVs mission, and solve many problems such as synchronous online change of mission, new UAV's joining in the formation, or one UAV's failure to complete mission due to breakdown and need of dropping out of the formation.

Acknowledgment. Supported by "the Fundamental Research Funds for the Central Universities", NO. NP2017209.

Supported by "Aviation Science Funds", NO. 20165852052.

Supported by "Advanced research project of the Ministry of army equipment development in 13th Five-Year", NO. 30102080101.

References

1. McLain, T.W., Beard, R.W.: Coordination variables, coordination functions, and cooperative timing missions. J. Guidan. Control Dyn. **28**(1), 150–161 (2005)
2. Szczerba, R.J., Galkowski, P., Glickstein, I.S.: Robust algorithm for real-time route planning. IEEE Trans. Aerosp. Electron. Syst. **36**(3), 869–878 (2000)

3. Bauso, D., Giarre, L., Pesenti, R.: Multiple UAV cooperative path planning via neuro-dynamic programming. In: 43rd IEEE Conference on Decision and Control, pp. 1087–1092 (2004)
4. de Castro, L.N., Tinunis, J.: Artificial Immune Systems: A New Computational Intelligence Approach, pp. 77–88. Springer Press, British (2002)
5. Gu, D.W., Kamal, W.: A UAV waypoint generator. In: AIAA 1st Intelligent Systems Technical Conference, pp. 1–6. Chicago, IL, USA (2004)
6. Dyke Parunak, H.V., Purcell, M.: Digital pheromones for autonomous coordination of swarming UAV's **10**(2), 654–655 (2002)
7. Wang, P.K.C.: Navigation strategies for multiple autonomous mobile robots moving in formation. Robot Syst. **8**(2), 177–195 (1991)
8. ZhengQiang, Y.: US air force in the next 20 years the development of a small UAV roadmap. Def. Point **10**, 58–59 (2016)
9. Xiaowen, Y.: The new technology in the future five aircraft. Air **1**, 88–89 (2012)
10. Lau, S.Y., Naeem, W.: Cooperative tensegrity based formation control algorithm for a multi-aircraft system. In: American Control Conference, pp. 750–756 (2015)

Optimal Antenna Selection for TCP Throughput Over Practical Distributed Cognitive Radio Networks

Fa Liu[1(✉)], Honglin Zhao[2], and Jiliang Zhang[3]

[1] China Electronics Technology Group Corporation No.10 Research Institute, Chengdu, China
silent_0602@163.com

[2] Communication Research Center, Harbin Institute of Technology, Harbin, China
hlzhao@hit.edu.cn

[3] School of Information Science and Engineering, Lanzhou University, Lanzhou, China
zhangjiliang@lzu.edu.cn

Abstract. In a cognitive radio networks (CRNs), the available idle spatial domain resource has to be considered to achieve a high spectral efficiency. In this paper, taking the constraint on the occupied time slot and frequency band into account, an antenna selection strategy for the maximized transport-layer throughput (AS-MTT) is proposed for secondary user (SU) using multiple-input multiple-output (MIMO) antenna arrays. The maximum achievable cross-layer transmit control protocol (TCP) throughput over practical distributed (PDCRNs) is formulated as an optimization objective. Considering the combination of physical-layer antenna selection, modulation, coding scheme, and data link layer frame size, the searching function for AS-MTT is modeled as max-min optimization problems. Numerical results show that the proposed AS-MTT strategy is capable to improve the SU's TCP throughput.

Keywords: PDCRNs · Antenna selection · TCP throughput · AS-MTT

1 Introduction

The rapid development of cognitive radio networks (CRNs) has attracted due to academic attention and great effort over last decades [1]. The essential idea of the CRNs technology is to exploit the available idle allocated spectrum resources [2]. Based on the observation results about the wireless electromagnetic environment, the secondary users (SUs) with low priority in CRNs are able to orthogonally access the frequency channels which is not occupied by primary users (PUs) at that moment. Existing researches investigated how to carry out spectrum sensing and accessing strategies in the time and the frequency domains. However, the available spatial domain resource, which provides the degrees of freedom

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 530–540, 2019.
https://doi.org/10.1007/978-981-13-6264-4_64

simultaneously, is ignored so far. As a promising technology for single-radio frequency (RF) multiple-input multiple-output (MIMO) network [3], the antenna selection technologies were employed to provide spatial diversity and to increase the reliability of wireless networks [4].

Furthermore, the current research activities on the sensing effectiveness and spectrum utilization are mostly limited to satisfy the design criteria of quality of service (QoS) in the physical layer. Inevitably, the performance parameters in the upper layer, such as the transmit control protocol (TCP) throughput, are neglected. However, the TCP throughput supports the significant amount of Internet traffic including WWW (HTTP), file transfer (FTP), e-mail (SMTP), and so on [5].

Previously, a hybrid noncooperative MAC protocol designed for the maximum throughput is proposed to search the integration of the optimal spectrum sensing and access [6]. In [7], a multiple-input multiple-output (MIMO) cross-layer transmit-antenna selection algorithm is proposed to maximize the average data rate in data link layer (DLL). [8] presents an opportunistic spatial orthogonalization (OSO) method to optimize the physical-layer throughput, while guaranteeing PU's interference tolerance. However, these papers [6–8] only pay attentions to QoS's optimization in physical layer or data link layer. A cross-layer design method is established in [9] to maximize the TCP throughput with partially observable Markov decision process (POMDP). In [10], the minimum retransmission problem accused by the inaccurate spectrum sensing is considered to design a cross-layer approach between transport layer (TPL) and medium access control (MAC) layer. From [9,10], the improvement of QoS in the transport layer is relied on utilization of available spectrum resource. However, in case that PU occupies all the frequency bands, the guarantee of SU's performance is not considered. Furthermore, the interference caused by PU to SU is not sufficiently studied in [6–10].

In this paper, we propose an antenna selection strategy within the MIMO antennas to maximize the TCP throughput (AS-MTT) in practical distributed cognitive radio networks (PDCRNs). Based on the AS-MTT strategy, both the PU and the SU can utilize the different spatial channels to transmit the data packets on the same frequency band at the same time. In particular, according to the TCP Reno protocol, the maximization of SU's TCP throughput is formulated as the max-min problem within the satisfaction of PU's interference threshold. Moreover, the transmission between PU's transceiver and SU's transceiver is modeled as the Rayleigh fading channels.

This paper is organized as follows: Sect. 2 introduces the system models; Sect. 3 presents the AS-MTT strategy for the maximized TCP throughput; Sect. 4 presents the numerical and simulation results under the proposed strategy; and Sect. 5 concludes the paper.

2 System Model

As shown in Fig. 1, SU's AS-MTT strategy over PDCRNs is illustrated. Equipped with N_t antennas in SU's transmitter and N_r antennas in SU's receiver, k SU's transmitter antennas are chosen to maximize the TCP throughput. On

the one hand, considering the practical communication situation, SU's receiver inevitably suffers from the interference made by PU's transmitter. On the other hand, since PU with higher priority of frequency bands usage, the interference in PU's transmission is constrained to satisfy PU's interference threshold. In particular, the wireless communication environment is modeled as the Rayleigh fading channels over PDCRNs.

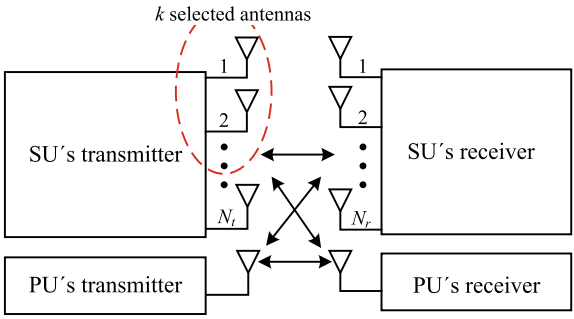


Fig. 1. SU's AS-MTT strategy in PDCRNs

2.1 Antenna Selection

Consider SU's transceiver with M_t transmit and M_r receive antennas in a Rayleigh fading channel. The signal received by SU's receiver is given by

$$y = \mathbf{H\Pi x} + \mathbf{n} \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^{N_t \times 1}$ are the signals transmitted by SU's transmitter; $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ and $\mathbf{\Pi} \in \mathbb{C}^{N_t \times N_t}$ are the i.i.d Rayleigh fading channel matrix and the permutation matrix dependent on the detection order, respectively [11]. $\mathbf{n} \sim N\left(0, N_0 \mathbf{I}_{N_r}\right)$ represents the circularly symmetric complex Gaussian noise vector, where $\mathbf{I}_{N_r}$ denotes the identity matrix of size N_r. Based on the assumption that the SUs know the channel state information (CSI), the QR decomposition for matrix $\mathbf{H}$ ($\mathbf{H\Pi} = QR$) using the zero forcing in SU's receiver is adopted to accomplish the decision feedback detector (DFD) [12]

$$y_R = \mathbf{QRx} + \mathbf{n} \tag{2}$$

where $\mathbf{Q} \in \mathbb{C}^{N_r \times N_t}$ is the interference suppression matrix with orthonormal columns. Meanwhile, $\mathbf{R} \in \mathbb{C}^{N_t \times N_t}$ is the upper triangular matrix with positive diagonal. Based on the k ($k \leq N_t$) antennas selected in SU's transmitter, the signals obtained by SU's receiver are given as

$$y_R^k = \mathbf{H}^k \Pi^k \mathbf{x}^\mathbf{k} + \mathbf{n}^k = \mathbf{Q}^k \mathbf{R}^k \mathbf{x}^k + \mathbf{n}^k \tag{3}$$

Assuming that the power sent by SU's transmitter is P_x, the SNR across jth sub-channel ($j \in \{1, 2, \ldots, N_t\}$) is expressed as

$$s_j = \frac{r_{jj}^2 P_x}{kN_0 \log_2 M} \tag{4}$$

where r_{jj} is the (j, j)th entry of $\mathbf{Q}$, and M is the modulation order.

Table 1. The BER of different modulations

Modulation	$P_b(s_j)$
BPSK	$P_b(s_j) = Q(\sqrt{2s_j})$
QPSK	$P_b(s_j) = Q(\sqrt{2s_j})$
MPSK	$P_b(s_j) \approx \frac{2}{log_2 M} Q\left[\sqrt{2s_j log_2 M} \sin[\frac{\pi}{M}]\right]$
M-QAM	$P_b(s_j) \approx \frac{4}{log_2 M} Q\left[\sqrt{\frac{3s_j log_2 M}{M-1}}\right]$

2.2 TCP Throughput

The approximate bit error rate (BER) using the coherent demodulations is given by [13].

The package A_z transmitted by SU in the TCP layer is separated into N_{fr} data frame Q_z at data link layer; meanwhile, the frame head containing the frame control information and MAC address is added to Q_z. Assume that the lengths of A_z and Q_z are L_{tcp} (bits) and L_h (bits), respectively. Therefore, the size of Q_z is denoted as

$$L_{fr} = \frac{L_{tcp}}{N_{fr}} + L_h \tag{5}$$

According to Table 1, the single data frame Q_z is transmitted in k parallel channels. Therefore, the probability of transmission failed in the jth sub-channel ($j \in \{1, 2, \ldots, N_t\}$) is given by

$$F_d^j = 1 - \left[1 - P_b(s_j)\right]^{\frac{L_{fr}}{k}} \tag{6}$$

Moreover, considering SU transmits the data in k parallel sub-channels, the failure probability of Q_z is written by

$$\mathrm{PER}(H^k, s_j) = 1 - \left[\prod_{j=1}^{k} (1 - F_d^j)\right] \tag{7}$$

Based on the retransmission of data in the TCP Reno protocol, SU's failure probability of single frame is denoted as

$$F_{\mathrm{Td}} = \left[\mathrm{PER}(H^k, s_j)\right]^{N_{\mathrm{rt}}+1} \tag{8}$$

Furthermore, according to the TCP Reno protocol, the failure probability of a TCP package received in SU's receiver is given by

$$P_{\mathrm{L}}^{\mathrm{TCP}} = 1 - (1 - F_{\mathrm{Td}})^{N_{fr}} \tag{9}$$

According to the automatic repeat request (ARQ) protocol, an ACK frame returns to SU's transmitter when every data frame is transmitted successfully. Therefore, the round-trip time (RTT) of TCP package is denoted as

$$\mathrm{RTT} = 2\mathrm{T_w} + \left[\frac{L_{fr}}{r}(N_{\mathrm{av}} + 1) + \frac{L_{\mathrm{ACK}}}{r}\right] N_{fr} \tag{10}$$

where $\mathrm{T_w}$ is the wired transmission delay. L_{fr} and L_{ACK} are the size of data frame and ACK frame, respectively. And r is the transmission rate of wireless link. N_{av} is the average transmission time for each data frame.

Moreover, assume that single SU's available channel band is B. Combining Table 1 and rate q turbo code, the transmission rate r with k selected antennas is given as

$$r = \frac{k \times q \times B \times \log_2 M}{2}(bits/s) \tag{11}$$

The long-term steady-state send rate of the TCP flow is characterized by the TCP throughput. Therefore, based on the TCP Reno protocol, the TCP throughput with the probability of packet loss P_L^{TCP} is given [5]

$$U(\mathrm{RTT},\, T_0,\, a,\, P_L^{\mathrm{TCP}}) \approx \frac{1}{\mathrm{RTT}\sqrt{\frac{2aP_L^{\mathrm{TCP}}}{3}} + T_0 \min\left(1, 3\sqrt{\frac{3aP_L^{\mathrm{TCP}}}{8}}\right) P_L^{\mathrm{TCP}}\left(1+32(P_L^{\mathrm{TCP}})^2\right)} \tag{12}$$

where T_0 is the initial value of the timeout, and a is the number of packages received before the acknowledgement arrives.

Considering the TCP throughput is limited by the maximum jam window K, and the throughput in the TCP layer actually is denoted as

$$U_T(K, \mathrm{RTT},\, T_0,\, a,\, P_L^{\mathrm{TCP}}) = \min\left[\frac{k}{\mathrm{RTT}}, U(\mathrm{RTT},\, T_0,\, a,\, P_L^{\mathrm{TCP}})\right] \tag{13}$$

Combining Eq. (12), the AS-MTT strategy without considering PU's interference limitation is modeled as P1 problem, when SU searches the available channel.

$$\mathrm{P1}: \underset{H^k}{\text{maximize}} \left[U_T(K, \mathrm{RTT},\, T_0,\, a,\, P_L^{\mathrm{TCP}}(H^k))\right] \tag{14}$$

Since the $U_T(K, \mathrm{RTT},\, T_0,\, a,\, P_L^{\mathrm{TCP}}(H^k))$ is constrained by the K, the P1 problem is simplified as max-min optimal problem $P1^\star$.

$$\begin{aligned} \mathrm{P1}^\star : &\underset{H^k}{\text{maximize}} \left[\min\left(\tfrac{K}{\mathrm{RTT}}, U_T(\mathrm{RTT},\, T_0,\, a,\, P_L^{\mathrm{TCP}}(H^k))\right)\right] \\ = &\begin{cases} \underset{H^k}{\text{maximize}} \left[U(\mathrm{RTT},\, T_0,\, a,\, P_L^{\mathrm{TCP}}(H^k))\right] & \text{if} \quad U(\mathrm{RTT},\, T_0,\, a,\, P_L^{\mathrm{TCP}}) < \frac{K}{\mathrm{RTT}} \\ \underset{H^k}{\text{maximize}} \left[\frac{K}{\mathrm{RTT}}\right] & \text{if} \quad U(\mathrm{RTT},\, T_0,\, a,\, P_L^{\mathrm{TCP}}) \geq \frac{K}{\mathrm{RTT}} \end{cases} \end{aligned} \tag{15}$$

3 AS-MTT Strategy for the Maximized TCP Throughput

Based on the "white space" in the time–spectrum domain, the cognitive radio has the opportunities to improve efficiently spectrum usage. However, the worst situation for SU cannot be avoided that there is no available spectrum resource at any time. In other words, PU's transceiver keeps in touch with each other all the time. Therefore, the way for SU's communication is exploited the spatial dimensions; meanwhile, the interferences made by SU's links are orthogonal to PU's link.

3.1 PU's Interference Limitation

Over the PDCRNs, we assume that PU's transceiver is equipped with one antenna. Therefore, according to Eqs. (1)–(3), the signal in PU's receiver is given by

$$y_{\mathrm{PU}} = h_{\mathrm{PU}} x_{\mathrm{PU}} + \sum_{n=1}^{k} h_{1,n} x_n + Z_{\mathrm{PU}} \tag{16}$$

where $Z_{PU} \sim N(0, N_0)$ is the additive white Gaussian noise (AWGN), and the channel gain h_{PU} is the Rayleigh fading channel coefficients between PU's transmitter and receiver [14]. Furthermore, the x_n and $h_{1,n}$ represent SU's signal sent by the nth transmit antenna and relevant coefficient of Rayleigh fading channel between SU's transmitter and PU's receiver, respectively. We define the PU's transmit power as P_{PU}. Using the opportunistic spatial orthogonalization [8], the SINR obtained in PU's receiver is written by

$$\mathrm{SINR}_{\mathrm{PU}} = \frac{\|h_{\mathrm{PU}}\|^2 P_{\mathrm{PU}}}{\sum_{n=1}^{k} \left[\frac{|\langle h_{\mathrm{PU}}, h_{1,n} \rangle|^2}{\|h_{\mathrm{PU}}\|^2} P_n \right] + N_0} = \frac{\|h_{\mathrm{PU}}\|^2 P_{\mathrm{PU}}}{\beta_k + N_0} \tag{17}$$

where $\beta_k = \sum_{n=1}^{k} \left[\frac{|\langle h_{\mathrm{PU}}, h_{1,n} \rangle|^2}{\|h_{\mathrm{PU}}\|^2} P_n \right]$ means the interference power, and P_n is the transmit power of SU's nth antenna. According to the PU's bandwidth B_{PU}, the PU's channel capacity without SU's interference is denoted as

$$c_0(k) = B_{\mathrm{PU}} \times \log_2\left[1 + \mathrm{SNR}_{\mathrm{PU}}\right] \quad = B_{\mathrm{PU}} \times \log_2\left[1 + \frac{\|h_{\mathrm{PU}}\|^2 P_{\mathrm{PU}}}{N_0}\right] \tag{18}$$

However, the interference made by SU's on the PU's receiver is not evitable in PDCRNs. Therefore, the PU's channel capacity under SU's interference is given by

$$c_1(k) = B_{\mathrm{PU}} \times \log_2\left[1 + \mathrm{SINR}_{\mathrm{PU}}\right] \tag{19}$$

Furthermore, the normalized maximum interference limitation on channel capacity c_{PU} under $\beta_k^{\max}$ is defined as

$$c_{\mathrm{PU}} = \frac{c_0 - c_1^{\max}(k)}{c_0} \quad = 1 - \frac{B_{\mathrm{PU}} \left\{ \log_2\left[\beta_k^{\max} + N_0 + \|h_{\mathrm{PU}}\|^2 P_{\mathrm{PU}}\right] - \log_2\left[\beta_k^{\max} + N_0\right] \right\}}{c_0} \tag{20}$$

where $\beta_k^{\max}$ is the maximum interference threshold tolerated by PU, when SU transmits data with k antennas. From Eq. (20), the $\beta_k^{\max}$ is given by

$$\beta_k^{\max} = \frac{\|h_{\mathrm{PU}}\|^2 P_{\mathrm{PU}}}{2^{\frac{c_0(1-c_{\mathrm{PU}})}{B_{\mathrm{PU}}}} - 1} - N_0 \tag{21}$$

Assume that SU knows PU's interference threshold $\beta_k^{\max}$ in advance. When the percentage of decreased PU's channel capacity is more than c_{PU}, SU stops to transmit data. In other words, by virtue of AS-MTT strategy, SU is able to select k transmit antennas to satisfy that sum of interference power β_k shadowed on PU's link is less than $\beta_k^{\max}$.

3.2 AS-MTT Strategy to Maximize the TCP Throughput

To access the available spatial resource, SU's transmitter selects k antennas to build the communication link. Therefore, the signals at SU's qth antenna are written by

$$y_q = \sum_{n=1}^{k} h_{q,i} x_i + h_{q,\mathrm{PU}} x_{\mathrm{PU}} + z_1 \tag{22}$$

where x_i is the signal sent by SU's ith antenna, $h_{q,i}$ ($h_{q,i} \in H_{N_r \times k}$) denotes the Rayleigh fading channel coefficient between ith transmitter and qth receiver [14]. The mean and variance of $h_{q,i}$ are $\sigma_{q,i}\sqrt{\frac{\pi}{2}}$ and $(2 - \frac{\pi}{2})\sigma_{q,i}^2$, respectively. Meanwhile, the $h_{q,\mathrm{PU}}$ denotes PU's signals arrives at SU's qth received antenna. z_1 means the AWGN with $z_1 \sim N(0, N_1)$. Furthermore, using the QR decomposition for matrix $H_{N_r \times k}$, SU's SINR in the mth ($1 \le m \le k$) sub-channel is

$$s_m = \frac{r_{mm}^2 P_x}{k\left[\sum_{q=1}^{N_r} \frac{|\langle h_{q,\mathrm{PU}}, h_{q,m}\rangle|^2}{\|h_{q,m}\|^2} P_{\mathrm{PU}} + N_1\right]} \tag{23}$$

Therefore, combining Eqs. (13)–(15) and Eq. (21), the AS-MTT strategy in SU's transmitter with the TCP reno protocol is modeled as the P2 problem.

$$\begin{aligned} \mathrm{P2}: \max_{\mathrm{H}_{N_r \times k}} & \left[\min\left\{\frac{K}{\mathrm{RTT}}, U\left[\mathrm{R}TT, T_0, a, P_{\mathrm{L}}^{\mathrm{TCP}}(\mathbf{H}_{N_r \times k})\right]\right\}\right] \\ \text{subject to } & \beta_k \le \beta_k^{\max} \end{aligned} \tag{24}$$

4 Simulation Results

In this section, we present the TCP throughput of SU's transmission with the AS-MTT strategy over PDCRNs. Simulation parameters are shown in Table 2.

Figure 2 demonstrates the TCP throughputs U using various modulations based on the AS-MPTT strategy. In Fig. 2, we notice that U first rises and

Table 2. Simulation parameters

B	Channel bandwidth	1 MHz
N_t	Number of SU's transmit antenna	4
N_r	Number of SU's receiver antenna	4
N_{fr}	Number of data frame	10
L_{tcp}	Length of A_z	12000 bits
L_h	Length of Q_z	20 bits
N_{rt}	Maximum retransmission time	5
T_w	Delay of TCP packets	15 ms
L_{ACK}	Length of ACK frame	24 bits
T_0	Initial timeout	2 s
K	Maximum congestion window	2500 bytes
q	Coding efficiency	3/4
$\sigma^2_{q,i}$	Variance of independent Gaussian distribution	$\frac{1}{10\sqrt{2}}$

then falls with P_x gradually increasing. Considering SU knows the PU's maximum interference threshold in advance, SU obtains the TCP throughput under the limitation of PU's tolerant threshold. Furthermore, according to Eqs. (20)–(21), SU's transmitter stops to send signals when the interference power exceeds the $\beta_k^{\max}$. In particular, SU's TCP throughput using high-order modulation efficiently improves with P_x increasing.

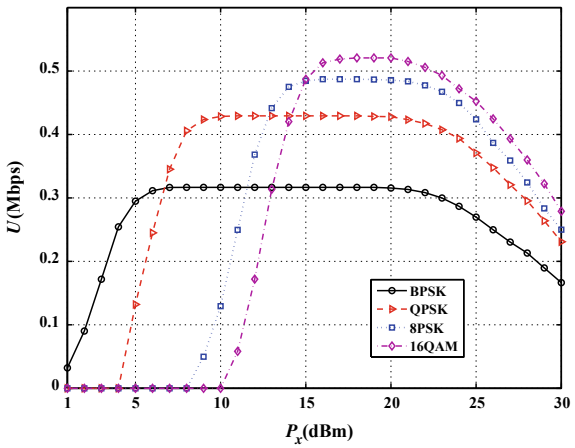


Fig. 2. Comparison of $\boldsymbol{U}$ with different modulations

Figure 3 shows the performance comparison between AS-MPTT strategy and S-strategy, where S-strategy means SU randomly chooses the transmit antenna.

In Fig. 3, the TCP throughput U with AS-MPTT strategy is higher than the U with S-strategy. The main reason is that the optimal transmit antennas are chosen by the AS-MPTT strategy without interfering with PU's transmission. Oppositely, S-strategy is not able to guarantee TCP throughput since the chosen antenna by SU easily makes interference to PU.

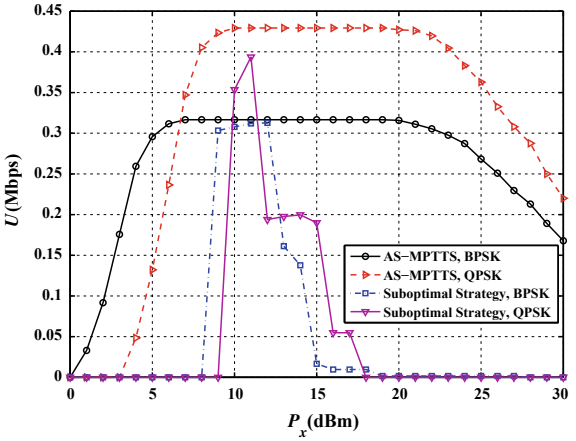


Fig. 3. Comparison of $\boldsymbol{U}$ based on AS-MPTT strategy and S-strategy

From Fig. 4, we observe that comparison of TCP throughput U with the different numbers of selected antenna. Figure 4 illustrates that comparing the single antenna, the increasing number of chosen antennas leads to higher TCP throughput U based on the same modulations.

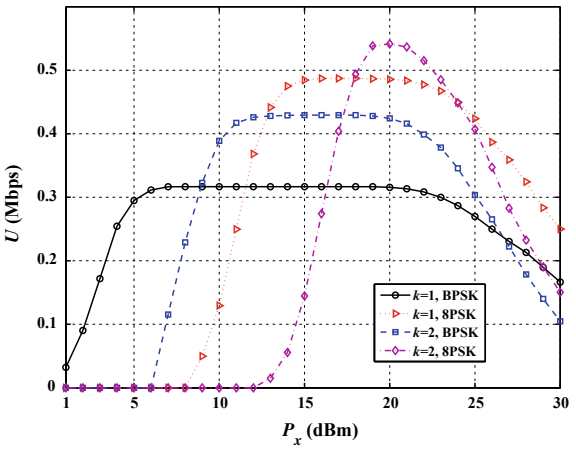


Fig. 4. Comparison of $\boldsymbol{U}$ based on different numbers of transmitted antennas

Figure 5 shows the comparison of different TCP throughputs based on the various PU's transmit powers, where the number of selected antenna k is 1. In Fig. 5, with PU's transmit power increasing, SU's TCP throughput also improves. The reason is that PU with higher transmitted power can tolerate more interference made by SU.

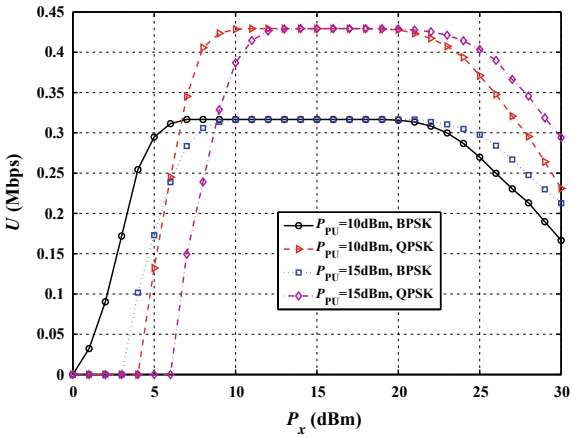


Fig. 5. Comparison of $\boldsymbol{U}$ based on different PU's transmitted powers

In Fig. 6, we observe the comparison of SU's performance based on the different PU's normalized maximum interference limitations c_{PU}. According to the higher PU's tolerance c_{PU}, the more TCP throughput is obtained.

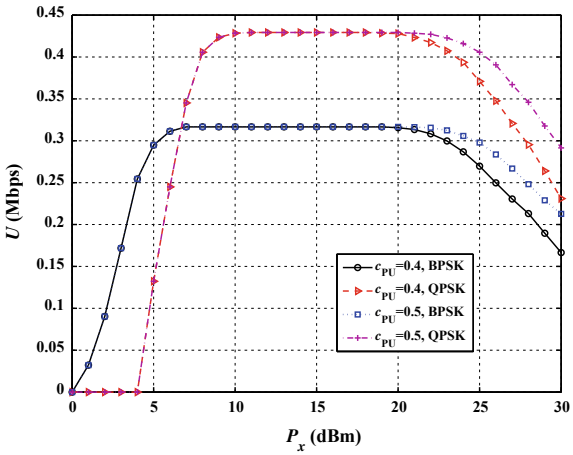


Fig. 6. Comparison of $\boldsymbol{U}$ based on different PU's interference limitations

5 Conclusion

In this paper, the AS-MTT strategy for SU equipped with MIMO antennas over PDCRNs is investigated. The available antennas for SU are employed to maximize TCP throughput. The problem is formulated as a max-min optimization problem subject to the PU's limitation of interference tolerance. The modulations and encoding in the PHY, the size and format of each frame in the DDL, and the TCP Reno protocol in the TPL are jointly considered. Relying on the AS-MTT strategy, SU's TCP throughput is guaranteed no matter when PU accesses the frequency bands.

References

1. Liu, F., Ma, Y., Zhao, H., et al.: Evolution handoff strategy for real-time video transmission over practical cognitive radio networks. China Commun. **2**, 141–154 (2015)
2. Mitola, J.: Cognitive radio architecture evolution. Proc. IEEE **97**, 626–641 (2009)
3. Zhang, J., Li, Q., Kim, K.J., et al.: On the performance of full-duplex two-way relay channels with spatial modulation. IEEE Trans. Commun. **64**(12), 4966–4982 (2016)
4. Amarasuriya, G., Tellambura, C., Ardakani, M.: Two-way amplifyand-forward multiple-input multiple-output relay networks with antenna selection. IEEE J. Sel. Areas Commun. **30**(8), 1513–1529 (2012)
5. Padhye, J., Firoiu, V., Towsley, D.F., et al.: Modeling TCP reno performance: a simple model and its empirical validation. IEEE/ACM Trans. Netw. **8**(2), 133–145 (2000)
6. Liu, F., Zhao, H., Zhang, C.: HNC-MAC: hybrid non-cooperative MAC protocol for independent secondary user over cognitive radio networks. In: Proceedings of IEEE WPMC 2014, pp. 429–433 (2014)
7. Ghosh, A., Hamouda, W.: Cross-layer design for cognitive MIMO ad-hoc networks. In: Proceedings of IEEE ICC, pp. 1–5 (2011)
8. Shen, C., Fitz, M.P.: Opportunistic spatial orthogonalization and its application in fading cognitive radio networks. IEEE J. Sel. Top. Signal Process. **5**(1), 182–189 (2011)
9. Luo, C., Yu, F.R., Ji, H., et al.: Cross-layer design for TCP performance improvement in cognitive radio networks. IEEE Trans. Veh. Technol. **59**(5), 2485–2495 (2010)
10. Xue, Y., Zhou, L.: TCP enhancement technology in cognitive network based on cross-layer designing. In IEEE ICUWB 2016, pp. 1–4 (2016)
11. Jiang, Y., Varanasi, M.K.: The effect of ordered detection and antenna selection on diversity gain of decision feedback detector. In: IEEE ICC 2007, pp. 5383–5388 (2007)
12. Varanasi, M., Guess, T.: Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel. In: Proceedings of the Thirty-First Asilomar Conference on Signals, Systems and Computers, pp. 1405–1409 (1997)
13. Goldsmith, A.: Wireless Communications. Press of the University of Cambridge, Cambridge (2005)
14. Proakis, J.G.: Digital Communications, 4th edn. Publishing House of Electronics Industry, Beijing (2006)

Research on Optimization of Public Space Addressing in Urban Open Block Based on Topological Complexity

Fu Erkang[1], Zhang Ping[1(✉)], Zhang Yujia[1], and Li Xinyun[2]

[1] Sichuan Agricultural University, Institute of Landscape Architecture, Room 605, Building 5,211 Huimin Road, Wenjiang District, Chengdu 611130, Sichuan, China
861804966@qq.com

[2] Sichuan Agricultural University, Institute of Landscape Architecture, Room 606, Building 5,211 Huimin Road, Wenjiang District, Chengdu 611130, Sichuan, China

Abstract. Based on the concept of topological complexity, this paper takes the Yulin District of Chengdu as an example. Based on a comprehensive consideration of the population and the distribution of 5G base stations, it starts with the selection of public spaces in urban open blocks to obtain the best public space addressing strategy for 5G user experience, and also conducted an in-depth study on the optimization of the addressing performance of the three-level public spaces in the urban open block: "block level-group level-neighbor level".

Keywords: Topological complexity · Open block · 5G user experience · Public space · Addressing performance

1 Introduction

The open block is a multi-functional compound city block characterized by high-density road network, open public landscape environment, excellent coverage of wireless signals such as 5G, and compact residential spatial structure [1]. Since 2016, <Several Opinions on Further Strengthening the Management of Urban Planning, Construction and Construction> has proposed the "Promotion of Block District System" by Central Committee [2], the research on open blocks has been increasing. Open blocks can share public spaces and create vibrant urban atmosphere. However, the process of rapid urbanization has broken the traditional public life pattern of traditional open blocks, resulting in the fragmentation of its public space and the decline of its vitality, which has led to a decline in the addressing performance of people within its space. Therefore, it is necessary to seek a scientific and effective way to optimize the addressing performance of public spaces in open blocks and restore their proper space dynamics.

In the future of 5G communications, wireless communication networks are evolving in the direction of network diversification, broadband, integration, and intelligence. With the popularity of various smart terminals, data traffic will grow in a

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 541–553, 2019.
https://doi.org/10.1007/978-981-13-6264-4_65

blowout manner. Future data services will be mainly distributed in indoor and hot spots. Open spaces in urban public spaces will be the core focus of data services. This makes the ultradense network become one of the main means to achieve 1000 times of future 5G traffic demand. The increasingly intensive network deployment also makes the network topology more complex. Peripheral area interference has become the main factor that restricts the growth of system capacity, greatly reducing the network energy efficiency, these have become important references indicator for the selection of public spaces in urban open blocks, and is gradually being integrated into the public space quality evaluation system.

2 Topological Complexity and Spatial Addressing Optimization

2.1 Topological Complexity

Although the topology complexity contains the word "topology", it is different from the word "topology" based on point set theory and logic from mathematicians. The topological complexity referred to in this paper belongs to the category of environmental psychology and was first proposed by MJ. O'neil (1991) based on the architectural layout. O'neil pointed out that there are some key positions (selection points) in the building interior. The addresser needs to make choices in more than two directions in order to decide on the direction to proceed. O'neil' called the method of calculating the complexity of architectural plane topology as the "interconnection density" (ICD) Measurement, ICD value represents the number of accessible paths between the various selected points in the building. The specific method is as follows: First, calculate the number of other points connected by each selection point, obtain the ICD value of each point, then add the ICD values of each point, and divide by the number of selected points to obtain the ICD value for each floor of the building [3] (Fig. 1).

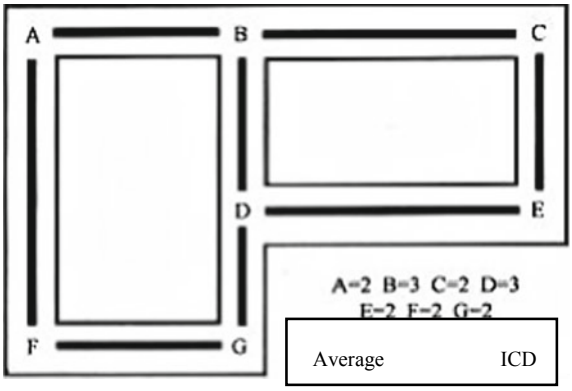


Fig. 1. The example of ICD value calculation in architectural plane Image Source: Hu Zhengfan, Lin Yulian《Environmental Psychology》

2.2 Spatial Addressing Optimization

When people address in space, the topological network in the space is a reasonable route for addressing. Through the ICD measurement, the addressable performance value of the addresser can be obtained thereby judging whether the supply and reachability of the public space node can meet people's addressing requirements in space. However, as people's enjoyment of public space is affected by both the spatial distribution and the quality of the destination space [4], merely increasing the supply of public space nodes and ensuring their accessibility cannot directly improve the spatial addressing performance. Therefore, the research on improving the addressing performance of public space should not only include the spatial ICD but also the quality of the space, and it is necessary to study the matching degree between them. When the matching degree between the two is optimal, the spatial addressing is optimal.

3 Optimization of Addressing of Public Spaces in Open Blocks

3.1 Overview of Public Space in Open Block

The public space in open block is an outdoor space used by the residents for daily life and social life in the block. It consists of buildings, roads, squares, green spaces, and ground environment facilities. It has "block level–group level–neighborhood level" transition. Among them, block level public spaces mainly serve the residents and visitors of this block, and the scale and openness of this level is the highest among the three levels, satisfying people's demand for large-scale public spaces. The group level public spaces are green parks, small landscape plazas, and pocket parks for residential groups in the residential group. Their service targets are mainly for the residents of this residential group, and the degree of openness is relatively high. Neighborhood level public spaces include small courtyards between small units in residential areas, and small spaces formed by visual gaps opened at frontages of residential areas. They mainly provide communication and negotiation for households or relatives, friends and passers-by of small units in the settlement. Places have a strong sense of realm.

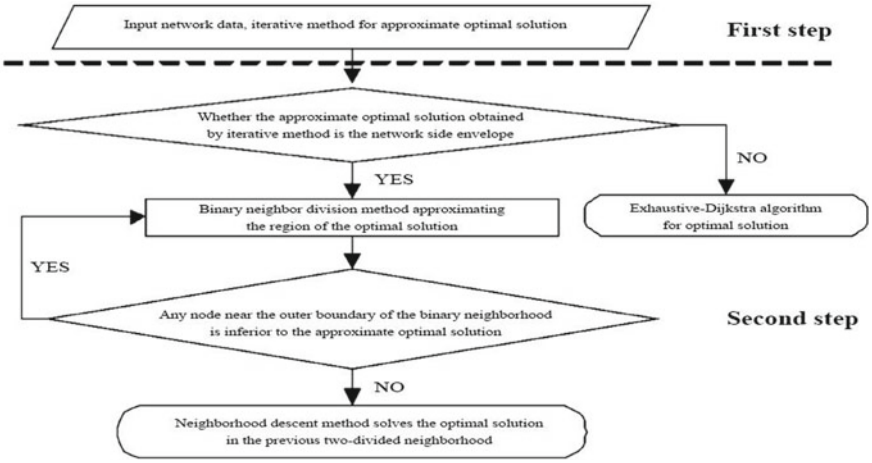
3.2 Addressing Optimization of Public Space in Open Block

According to the discussion of 1.2 above, the addressing performance of public space in open block should be measured by the degree of "spatial quality-ICD" matching of the main node space.

3.2.1 Optimum Modeling of Public Space Addressing in Open Blocks

The addressing optimization model based on topologically complex network clustering method is divided into two steps. The first step is to obtain the approximate optimal solution. The model uses iterative method to obtain the approximate optimal solution [5]. The second step is to approximate the optimal solution. The model uses a two-point

neighborhood descent method to gradually approximate the optimal solution. The model flow chart is shown below.



Assume that select a point in the network as a urban open block selection repository. The site selection point must be located on a certain network node, and the sum of the weighted shortest paths to the nodes of the network should be the smallest. The practical significance of this site selection is that when the nodes of the network are in a saturated state of people flow at the same time, the overall quality of the selected block address is the best.

The following variables and formulas are defined: Pi is a known network node (xi, yi) (i = 0, 1, 2, … n). P is the most preferred site (x, y) to be solved, which is the optimal solution of this case. Ci is the integrated spatial quality at the network node Pi. b is the space quality demand for the node's traffic. Assuming that each node in the network is in the saturation state of the human flow at the same time, the definition di is the total space quality demand of the project-opening block at the most optimal site P to Pi. Di is a weighted distance expressed by Eq. (1):

$$\mathrm{di} = \mathrm{b} \cdot \mathrm{ci} \cdot (\mathrm{x} - \mathrm{xi})2 + (\mathrm{y} - \mathrm{yi})2 \tag{1}$$

di can also be considered as the network traffic at the optimal location points P to Pi. Assuming that the network has n nodes, the sum of the overall spatial qualities of all nodes to the most preferred site P is D, expressed by Eq. (2):

$$D = \sum_{i=1}^{n} d_i = \sum_{i=1}^{n} b \cdot c_i \cdot \sqrt{(x - x_i)^2 + (y - y_i)^2} \tag{2}$$

Using Newton iteration method to solve D:

$$\begin{cases} \frac{\partial D}{\partial x} = \sum_{i=1}^{n} \frac{b \cdot c_i \cdot (x - x_i)}{\sqrt{(x-x_i)^2 + (y-y_i)^2}} = 0 \\ \frac{\partial D}{\partial y} = \sum_{i=1}^{n} \frac{b \cdot c_i \cdot (y - y_i)}{\sqrt{(x-x_i)^2 + (y-y_i)^2}} = 0 \end{cases} \tag{3}$$

If use a vector notation T = (x, y) T, F = (∂ D/∂ x, ∂ D/∂ y) T. The equation set (3) can be written as:

$$F(x, y) = 0 \tag{4}$$

$$F'(x, y) = \begin{bmatrix} \frac{\partial^2 D}{\partial x^2} & \frac{\partial^2 D}{\partial x \partial y} \\ \frac{\partial^2 D}{\partial y^2} & \frac{\partial^2 D}{\partial x \partial y} \end{bmatrix} \tag{5}$$

among

$$\begin{aligned} \frac{\partial^2 D}{\partial x^2} &= \sum_{i=1}^{n} \left[\frac{b \cdot c_i}{\sqrt{(x - x_i)^2 + (y - y_i)^2}} - \frac{b \cdot c_i \cdot (x - x_i)^2}{\sqrt{\left[(x - x_i)^2 + (y - y_i)^2\right]^3}} \right] \\ &= \frac{b \cdot c_i \cdot (y - y_i)^2}{\sqrt{\left[(x - x_i)^2 + (y - y_i)^2\right]^3}} \end{aligned} \tag{7}$$

$$\begin{aligned} \frac{\partial^2 D}{\partial y^2} &= \sum_{i=1}^{n} \left[\frac{b \cdot c_i}{\sqrt{(x - x_i)^2 + (y - y_i)^2}} - \frac{b \cdot c_i \cdot (y - y_i)^2}{\sqrt{\left[(x - x_i)^2 + (y - y_i)^2\right]^3}} \right] \\ &= \frac{b \cdot c_i \cdot (x - x_i)^2}{\sqrt{\left[(x - x_i)^2 + (y - y_i)^2\right]^3}} \end{aligned} \tag{8}$$

$$\frac{\partial^2 D}{\partial x \partial y} = -\sum_{i=1}^{n} \frac{b \cdot c_i \cdot (x - x_i)(y - y_i)}{\sqrt{\left[(x - x_i)^2 + (y - y_i)^2\right]^3}} \tag{9}$$

we take it as (x, y)k + 1 we get:

$$(x, y)^{k+1} = (x, y)^k - F'\left((x, y)^k\right)^{-1} F\left((x, y)^k\right) \tag{6}$$

Given the fact that (xi, yi)i = 1, 2 … n is known, the Newton iterative method is a quick method. For this type of problem, any point in the definition domain of the initial value selection is not sensitive, and the root of the Newton method is quadratically

converged. However, the formula of Newton's iteration method is relatively complex, and second-order partial derivatives are required, and the inverse of the matrix is also required. Then, the two-dimension neighborhood division method is used to reduce the optimal solution candidate address, and the descending method is used to gradually approximate the theoretical optimal address.

3.2.2 The Spatial Quality of Public Space in Open Block Public

The spatial quality of open blocks is the comfort level of people's activities in different levels of public space. According to environmental psychology and human needs, we divide the influencing factors into two aspects: (1) visual experience such as identification system, characteristic landscape, and good visual field; (2) Other sensory experiences such as noise pollution, gas contamination, dust, roof drip. Based on the "public space quality standard" (also known as PSPL) [6] proposed by famous Danish urban designer Jan Gael in public space research, this paper applies it to the quality evaluation of open block public spaces and screens them by means of expert scoring. Give a reasonable factor and assign weights to the factors (Table 1).

Multi-element Analysis Formula

Table 1. Open block public space quality evaluation table

Category	Factor	Weight
Visual experience	Good identification system	1
	Characteristic building facade	0.8
	Characteristic ground pavement	0.8
	Arbor and shrub	0.7
	Concentrated landscape elements: green areas and water	0.7
	Landmark	0.8
	Neat and orderly	0.8
	Good view	0.7
	Good detail	0.6
	Comfort materials	0.6
	High-quality lighting(Especially at night)	1
	Sense of place/The sense of theme and sense of atmosphere	0.4
Other sensory experiences	No noise	1
	No gas pollution	1
	No heavy dust	0.8
	No roof drip	0.8
	Wireless signal interference is small	0.8
	5G and other wireless network coverage	0.9
	Will not face cold\hot(windshield)	0.7
	Different activities could be carried out during the day and night	0.6
	Have borders to rest and park	0.7
	Could talk to others	0.4
	Meet the human scale comfort	0.8

4 Investigation and Analysis of Addressing Performance of Public Space in Yulin District

4.1 Overview of Yulin District Public Space

The research selected a case is located in the northeastern part of Wuhou District of Chengdu City. It is a traditional open-plan neighborhood with an early form of Chengdu and a "small block" shape. The research area is enclosed by Yulin Street, Yulin East Road, Yulin Middle Road and Nijiaqiao Road. It covers a total area of approximately 18.52 hectares and is a material carrier for the old Chengdu lifestyle.

Through literature review and on-site survey, we sorted out the current status of public space in Yulin District according to the spatial level of "block level-group level-neighbor level", and selected 16 public space nodes at the block level. 16 public space node at the group level, and 13 nodes at the neighborhood level (Fig. 2).

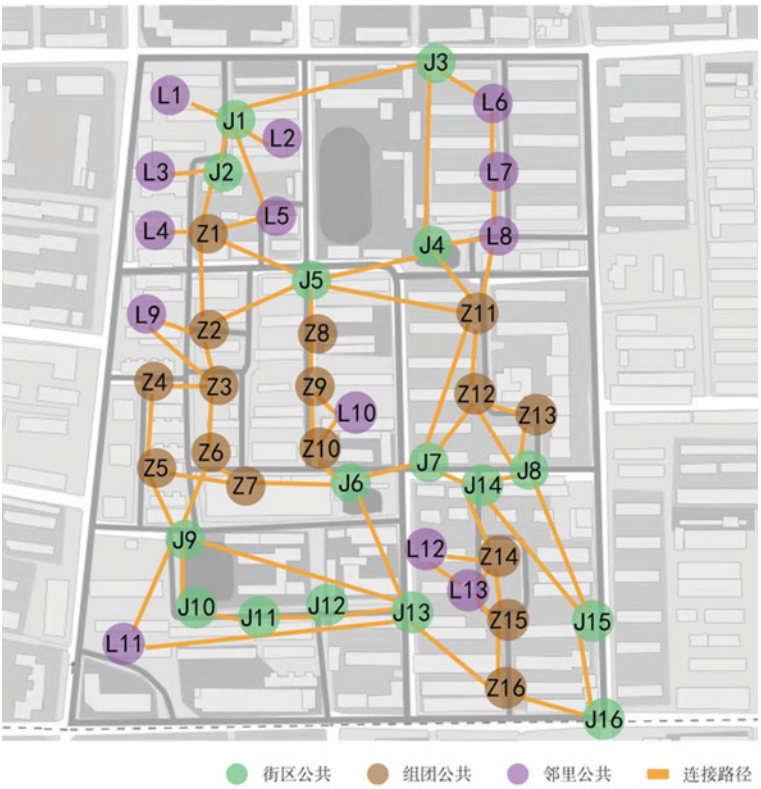


Fig. 2. Yulin District public space nodes and contact routes Image Source: Painted by author

4.2 ICD of Public Spaces in Yulin District

According to the site survey, there are 134 contact paths in the main nodes of public space in Yulin Street (Fig. 2). Degree (or connection) ki is the simplest and most important concept for characterizing and measuring a node. It means the quantity of edges the point i has [7]. In other words, the greater the degree of a node, the more important it is in the network. It can be defined according to its adjacency matrix [8]:

$$k_i = \sum_{j \in N} e_{ij}$$

The probability that the degree is greater than k using the degree of accumulation formula is as follows:

$$P_k(K > k) = \sum_{k' > k}^{\infty} p(k').$$

The intervenient Bi of the alternative address i can be expressed as:

$$B_i = \sum_{i,j \in N, j \neq k} \frac{n_{jk}(i)}{n_{jk}},$$

Among them, n_{jk} is the number of shortest paths connecting the points j and k, $n_{jk}(i)$ is the number of shortest paths connecting j and k and passing point i [9]. The ICD values for each node are shown in Table 2. The total public space in Yulin District is ICD = 134/45 = 2.98.

Table 2. "Space quality-ICD" match degree of Public space in Yulin District

The level of space	Node name	Space quality	ICD	Space quality/ICD
Block level	J1	2.14	5	0.43
	J2	2.03	3	0.68
	J3	1.76	3	0.59
	J4	3.21	4	0.80
	J5	2.76	5	0.55
	J6	3.26	4	0.81
	J7	3.74	4	0.93
	J8	1.66	4	0.41
	J9	1.36	5	0.27
	J10	1.39	2	0.69
	J11	2.41	2	1.20
	J12	2.03	2	1.02
	J13	2.80	5	0.56
	J14	2.64	4	0.66
	J15	3.29	3	1.10
	J16	3.28	2	1.64

(*continued*)

Table 2. (*continued*)

The level of space	Node name	Space quality	ICD	Space quality/ICD
Group level	Z1	1.73	5	0.35
	Z2	1.83	4	0.46
	Z3	3.18	4	0.80
	Z4	1.90	2	0.95
	Z5	1.91	3	0.64
	Z6	1.66	3	0.55
	Z7	1.18	3	0.39
	Z8	3.28	2	1.64
	Z9	2.58	3	0.86
	Z10	1.59	3	0.53
	Z11	2.98	5	0.60
	Z12	1.59	4	0.40
	Z13	2.38	2	1.19
	Z14	3.01	4	0.75
	Z15	1.65	3	0.56
	Z16	1.90	3	0.63
Neighborhood level	L1	2.72	1	2.72
	L2	2.72	1	2.72
	L3	2.25	1	2.25
	L4	1.98	1	1.98
	L5	2.56	2	1.28
	L6	2.40	2	1.20
	L7	2.30	2	1.15
	L8	2.12	3	0.71
	L9	2.63	2	1.32
	L10	2.42	2	1.21
	L11	1.02	2	0.51
	L12	2.01	2	1
	L13	1.91	3	0.64

Basic data sources: Investigation on addressing performance of public space in Yulin District

4.3 Analysis on the Space Quality of Public Space in Yulin District

The space quality of the main nodes in Yulin District public space can be investigated according to the "Open Block Public Space Quality Evaluation Table" in Table 2 above. In the visual experience, six factors including "good signage system, distinctive building facade, distinctive ground paving, arbor and shrub, and concentrated green space water body landscape elements and landmarks" were used by the investigator to investigate the nodes Space Center, according to the "viewing object and people's line of sight angle of 45° (vertical viewing angle) to watch the details, and people within 10 m

of viewing experience will become interesting and exciting" as the standard of judgment, the best visual experience Score points for scores of 7–10, generally 3–6, and not strong 1–2; other factors in the visual experience and all factors in other sensory experiences are highly consistent with these factors is 8–10, conform to the factors is 4–7, do not meet the factors is 1–3, highly inconsistent with these factors is 0 points [10]. By summing the scores of each node's space, we have obtained the evaluation table of the spatial quality of the main nodes of public space in Yulin District (Table 2).

4.4 Analysis on Matching Degree of "Space Quality-ICD" of Major Space Nodes in Yulin District

4.4.1 Analytical Method

Through the calculation of the ratio of ICD and space quality of 45 major space nodes in Yulin District, the "space quality-ICD" match degree was judged. The ratio between the two is high, which means that the node has high spatial reachability, but the space quality is low. It is necessary to improve its space quality to enhance the matching of its "space quality-ICD". The ratio between the two is low, which means that the space quality of this node is high, but the accessibility is poor. It is necessary to improve its reachability to enhance the matching of its "space quality-ICD". It should be noted that because people have different requirements for the accessibility and space quality of the public space at the level of block, group and neighborhood, so the analysis should use different criteria.

4.4.2 Analysis Results

(1) Analysis of Matching Degree of "Space Quality-ICD" in Block Level Public Space

There are 16 block level public space nodes. Based on the ratio of space quality score and ICD, the ratio lower than 0.5 is judged to be low space quality- high reachability node, and the ratio higher than 1.5 is the high space quality -low reachability node. It is concluded that there are 4 node spaces that need to be optimized (Fig. 3).

(2) Analysis of Matching Degree of "Space Quality-ICD" in Group Level Public Space

There are 16 group level public space nodes. Based on the ratio of space quality score and ICD, the node with the ratio lower than 0.75 is judged to be the low space quality —high reachability node, and the node with the ratio higher than 1.25 is judged to be the high space quality-low reachability node. It is concluded that there are 11 node spaces that need to be optimized (Fig. 4).

(3) Analysis of Matching Degree of "Space Quality-ICD" in Neighborhood Level Public Space

There are 13 neighborhood level public space nodes. Because of their spatial characteristics, the actual space quality is the main basis, and the ratio of the spatial quality score to the ICD is the secondary basis. The ratio below 1 is the low space quality node,

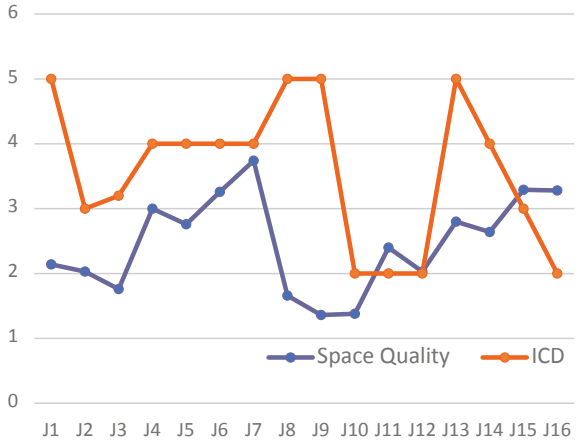


Fig. 3. Matching Degree of "Space Quality-ICD" in block level public Space Basic data sources: Investigation on addressing performance of public space in Yulin District

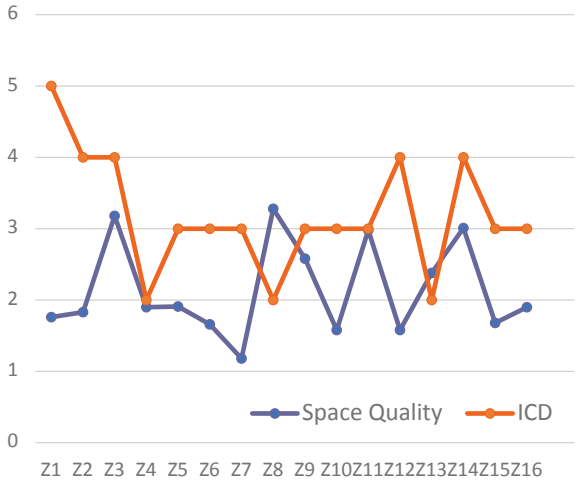


Fig. 4. Matching Degree of "Space Quality-ICD" in Group level Public Space Basic data sources: Investigation on addressing performance of public space in Yulin District

and the ratio is between 1–2 is medium space quality node. It is determined that the ratio is greater than or equal to 2 is high space quality node, and it is found that there are 4 node spaces that need to be optimized (Fig. 5).

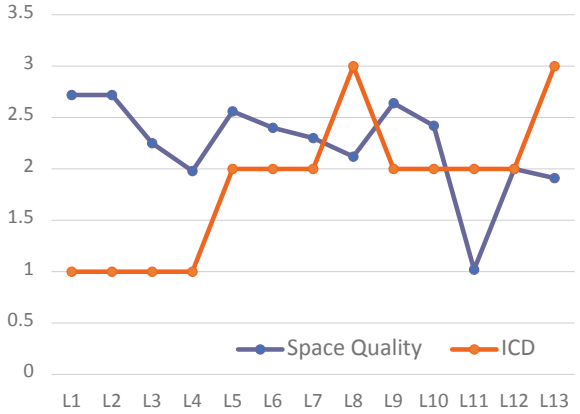


Fig. 5. Matching Degree of "Space Quality-ICD" in neighborhood level public space Basic data sources: Investigation on addressing performance of public space in Yulin District

5 Conclusion

We suggest that the space be classified into three levels of "block-group-neighborhood" for classification optimization, and the corresponding accessibility and spatial quality of the space at different levels are optimized respectively. Based on the concept of topological complexity, this research analyzes the matching degree of "ICD-spatial quality" in the main node space of open block public spaces. By optimizing the space of nodes at different levels, the goal of improving the addressing performance of public space in slice areas is achieved.

5.1 The Node of High ICD- Low Space Quality Space Should Be Optimized for the Space Quality

The high ICD node space is attracted by its high reach. By improving the spatial quality of this type of node space, the use of public spaces in open blocks is more comfortable and the addressing performance of people in their space is optimized.

5.2 The Node of Low ICD High Space Quality Space Should Be Optimized for Its Low Reach

The high-quality node space provides people with opportunities for public activities in open block, but if the ICD is low, this kind of space will be overlooked due to poor accessibility, resulting in a waste of space. By adding interesting space nodes to the path to this type of node space, its accessibility is enhanced and its spatial addressing performance is improved.

5.3 The Matching Degree of "ICD-Space Quality" in Public Space of Urban Open Block Should Be Optimum

When the matching degree of "ICD-space quality" of public spaces in urban open block is optimal, the site selection performance of the space nodes is optimal. In other words, public space nodes with high reachability should have good spatial quality.

References

1. Yuhang, S.: Research on the Opening Structural Pattern of the Urban Residential Block. Dalian University of Technology (2015)
2. Baojun, Y.: Discussion on open blocks. City Plan. Rev. **40**(12), 113–117 (2016)
3. Zhengfan, H., Yulian, L.: Environmental Psychology. China Building Industry Press (2012)
4. Frank, L., Devlin, A., Johnstone, S.: Neighbourhood design, travel, and health in Metro Vancouver: using a walkability index (2010)
5. Jicai, D., Huiming, Z.: The optimal location selection model based on complex network clustering. Sci. Geogr. Sin. **33**(2), 143–149 (2013)
6. Gehl, J., Gemzoe, L.: Public Spaces Public Life. China Building Industry Press (2003)
7. Yanlai, L.: Study on the Complexity of Logistics Network Structure and Optimization Design. Beijing Jiaotong University (2011)
8. Jianjun, W.: Study on Complexity of Topology in Urban Traffic Network. Beijing Jiaotong University (2008)
9. Ke, Q., Peng, Z., Xiangming, Y., et al.: Cluster analysis of urban rail transit nodes. Logist. Technol. (5), 92–95 (2014)
10. Jingjing, H.: Social Worker Occupational Psychology: Questionnaire Establishment and Characteristic Analysis. Nanjing Normal University (2014)

Routing Based on Community Division, Vitality, and Energy Balance in Delay Tolerant Networks

Dongbo Liu, Rongxi He(✉), and Cunqian Yu

College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China
hrx@dlmu.edu.cn

Abstract. With the popularity of mobile portable devices, massive social applications have been emerging and promoting the practice of social network, which provides a new opportunity for the advance of routing by applying social algorithm to Delay Tolerant Networks (DTNs). In this paper, we propose a novel routing protocol based on community division, vitality and energy balance (RCVE) for DTNs, which simultaneously considers the physical and social attributes of nodes. A new community division method is firstly presented in RCVE to divide various nodes into different communities according to the historically social information. Based on it, a vitality and energy balance based routing strategy is also proposed to further improve routing efficiency and to reduce network overhead. Finally, the performances of the proposed algorithm are verified through a series of simulation experiments. The results show that RCVE can increase the average delivery ratio, reduce the average overhead, and average hop compared with the existing algorithms.

Keywords: Delay Tolerant Networks · Routing · Community division · Vitality

1 Introduction

With the primary superiority to handle the communication problems in limited environment, e.g., frequent network disconnection, high latency, etc, Delay Tolerant Networks (DTNs) have been drawn attention from plenty of researchers. However, owing to the mobility and low node density in DTNs, it is hard to maintain a continuous end-to-end path in the rapid time-varying network topology, making data forwarding extremely difficult [1]. Furthermore, nodes in DTNs

This work was supported in part by the National Natural Science Foundation of China under Grant 61371091, the 13th five-year Key Research Project of DMU under Grant 3132016318, and the Fundamental Research Funds for Central Universities under Grant 3132017078.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 554–562, 2019.
https://doi.org/10.1007/978-981-13-6264-4_66

suffer from unpredictable movement and little knowledge of the network topology. As a result, how to design well-function routing protocols in DTNs with harsh conditions is an urgent front-burner problem.

Currently, massive social applications have been emerging and achieving the sharing of data from various wireless portable devices, effectively promoting the practice of social networks. The data produced from these multimedia services provide possibility for the applying of social algorithms and services, offering a new opportunity for the innovation of routing protocols in DTNs. These social networks attempt to analyze the social properties of these nodes in support of effectively promoting the design of efficient routing algorithms [2]. Among, community division algorithms have typically obtained quite a few attention from researchers due to its application to data forwarding [3]. Especially, the handheld mobile devices in DTNs are completely affected by the behavior patterns of carriers. So nodes can be analyzed and defined through the extracted certain social attributes. Traditional routing protocols only depend on geographical location and other physical attributes of nodes in DTNs. However, the routing strategy in society-based DTNs makes better use of the potential social attributes of nodes such as community structure [4]. By utilizing the discovered social attributes, routing performance can be effectively improved in DTNs. For example, people prefer to share messages with friends rather than strangers. The same cases exist in DTNs as well. When nodes forward massages in communities, the reliability of routing decision can be improved, the cooperation and data sharing among nodes can be promoted, and further the utilization rate of limited network resources can be advanced.

In this paper, we propose a routing protocol based on community division, vitality, and energy balance (RCVE), which not only takes advantage of the physical attributes, but also considers the social attributes of nodes. In RCVE, various nodes in DTNs are divided into different communities through utilizing the historically social information. Members in community are constantly updated based on friendship values between nodes, and the community is overlapped. On this basis, the routing strategy based on vitality selection and energy balance is introduced to improve routing efficiency and reduce network overhead. Simulation experiments based on the Opportunistic Networking Environment (ONE) simulator prove that our proposal increases the average delivery ratio, and reduces the average overhead and average hop compared with the existing algorithms.

The remainder of the paper is structured as follows. In Sect. 2, a brief review of related work is presented. Section 3 describes our routing algorithm in detail. Section 4 gives the simulation results. Finally, conclusions are drawn in Sect. 5.

2 Related Work

Our work mainly focuses on a routing based on community division, which devotes to the research of rational node community division and effective routing in DTNs. Therefore, work related to the presented paper will be summarized from the aspect of community division and routing algorithm.

The community division is commonly based on three structure criteria: topolo gy-based, membership-based and time-based. But the methods of dividing community have the varied forms. Lu et al. [5] exploit the conductance to detect the friendship among nodes for characterizing nodes in communities. In addition, they propose two efficient data forwarding metrics: intra-centrality and inter-centrality. Thus the shortest path within local community or between communities can be discovered. Zhang et al. [6] introduce a fast clique percolation algorithm Weak-CPM for overlapping community detection. Weak-CPM also characterizes the similarity between weak cliques, and thus the weak cliques can be judged to merge into a community, which decreases the calculated cost.

Generally, routing algorithms which do not consider the social attributes obtain high delivery ration at the expense of more energy and buffer space, such as Epidemic [7] and an improved probabilistic routing algorithm (PRTI) [8]. Many routing algorithms introduced in recent work improve routing performances by making use of the social attributes. Hui et al. [9] propose distributed routing algorithms to detect both static and temporal communities. Meanwhile, the accuracy to choose right paths in their routing based on simple community detection (RSCD) can be approximate up to 90% with suitable configuration. Li et al. [10] present a new two-stage spreading and forwarding dynamic routing algorithm to decrease the waste of resource. On the one hand, it is a multi-copy spreading strategy based on the social activity degree of nodes. On the other hand, it is a single-copy forwarding strategy based on the physical contact factor for portable device users between nodes. They also propose a new modeling method to describe a dynamic evolving process of the topology structure in DTNs. Zhou et al. [11] pick up on the moving tend on the basis of the static and dynamic feature of the subscribers. The dynamic community discovery is proved effective by the test using Twitter data.

3 Algorithm Description

3.1 Community Division Based on Friendship

A given network is defined as a graph $G = (V, E)$, where V is the set of vertices representing nodes or objects, and E is the set of edges representing their relationship. The friendship value $F(s, d)$ is defined to represent interactive relationship of edge between nodes s and d [12]. Communities C_i compose the membership-based network structure $G = \{C_1 \cup C_2 \cup ... \cup C_n\}$. Random movement of nodes in DTNs limits our attention to estimating $F(s, d)$ by using the contacted history between nodes s and d. In Fig. 1(c), there are several contacted records between s and d, which have clearly shown the important influencing factors: the contacted frequency, duration and regular. Comparing (1) with (2), they have the same time duration d_i of communication. Even so, their different contacted frequencies t_i lead to different friendship values. Comparing (2) with (3), in spite of the same regular, diverse durations produce varying results.

In this subsection, we propose a community division method through measuring and calculating friendship. Differing from previous works considering only

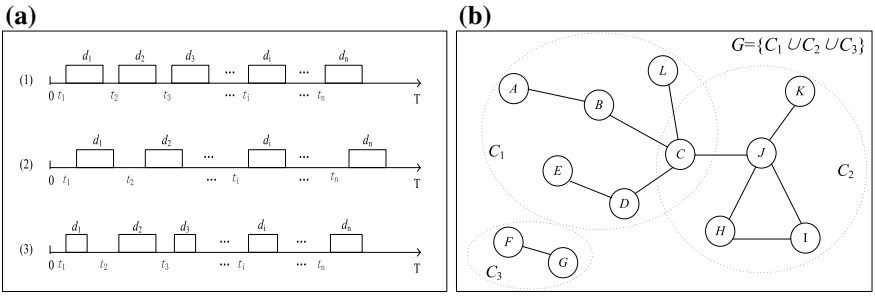


Fig. 1. (a) Different contacted records; (b) The communities: C_1, C_2, C_3.

one factor, we propose a new strategy to measure the friendship of the node pair. Not only the contacted frequency, but also the contacted duration and regular are all considered in our metric. Moreover, the validity of historical information and the indirect friendship are taken into account for the community division. In our method, each contacted records of the node pair are saved in contacted $List < Duration >$, which records friendship factors between nodes. When the friendship value of a node pair exceeds the predicted value, they will be screened and stored in the $List < familiarSet >$ for each other. For example, in Fig. 1(b), according to the historical records $List < Duration >$ and contact frequency in routing table of two nodes C and F, the results, $ListC = \{A, B, D, E, F, H, J, K, L\}$ and $ListF = \{G\}$ is obtained.

In the following, the process of our method is introduced in detail. The friendship value $F(A, B)$ is calculated and the familiar nodes which meet both two nodes A and B are selected through the following Eqs. (1–3). In Eqs. (1–3), n denotes the number of contact during the time interval [0,T], and $f(t)$ denotes the remaining time of the first contact after time t. F_n denotes the single friendship calculated for the contact in nth time. Since the contact between nodes randomly varies according to the location of the nodes, the contribution of the contacted history between nodes to the calculation of friendship values decreases with the increasing of time. Thereby, we defined α as the attenuation coefficient to modify the friendship value, for the reason that the contact regular is considered. α is an adaptive value with the increasing of time, which could improve the computational accuracy. t_i is the time difference between the beginning of the current communication and the end of the previous communication, and $t_{average}$ is the average value of t_i before time t. The representation of friendship value between node A and B is shown as follows:

$$F(A, B)_{n+1} = \alpha \times F(A, B)_n + (1 - \alpha) \times F_{n+1}. \tag{1}$$

$$F_n = \frac{2n^2 \int_{t_i}^{t_i + d_i} f(t)\, dt}{T^2}. \tag{2}$$

$$\alpha = 1 - \frac{t_i}{2 \times t_{average}}. \tag{3}$$

When the friendship value between node pair exceeds the specific threshold, two nodes will become a member in $List < familiarSet >$ for each other. Then the pair of nodes compare their $List < familiarSet >$, and the similarity is calculated according to Eq. 4. When the similarity is more than a certain degree, the pair of nodes will form a community. In the same way, more nodes will join in this community with the following Eq. 5. If the similarity $S(A, C_B)$ between node A and a community C_B is bigger than the half length of $ListA < familiarSet >$, this nodes will join in community C_B as well. Since the dynamic mobility characteristics of the nodes, the formed communities is overlapped, which can act as a bridge for communication between communities and enhance the data transmission rate among communities. And then communities which have $k-1$ nodes in common merge into a big community, which is similar with the K-Clique Percolation Method about Communities merging in [13]. Where $S(A, B)$ is the similarity of node A and B. What's more, the numerator denotes the number of common nodes between node A and B, and the denominator denotes the average length of $ListA$ and $ListB$. A greater similarity between nodes provides more communication chances.

$$S(A, B) = \frac{|ListA \cap ListB|}{\sqrt{|ListA| \times |ListB|}}. \tag{4}$$

$$S(A, C_B) = \frac{|ListA \cap ListC_B|}{|ListA| \times |ListC_B|}. \tag{5}$$

3.2 RCVE

In the DTNs, nodes carrying messages move randomly. Due to unknowing the accurate information of the destination node, nodes will choose a relay node that is closer to the destination nodes. However, the standards of the "closer" in deferent algorithms are much more diverse. For example, in Prophet, the greater contact probability is defined as "closer", and in the distance algorithm, the shortest distance is the decision standard. In our algorithm, the vitality and the energy level is both used as the decision standard.

On the basis of our community division, a routing strategy based on node vitality and energy consumption is proposed. This method avoids using the accidental encountered information between nodes that could result in wrong routing decision and waste of resources. According to the node vitality in the local community, when node C has a message m to forward, the relay node is dynamically selected. And the vitality of nodes in the community is regarded as a primary standard of "closer" in the RCVE. On one hand, the vitality of node A means the times and frequency that node A has communications with others. On the other hand, superabundant communications mean a great deal of energy consumption. Thus the node with second largest vitality in community is selected as the relay node. Meanwhile, to achieve energy balance, the residual energy of the relay node is not smaller than the medium level among all the nodes in local community. Otherwise, the algorithm will seek the relay node again. In Eq. 6,

$V(t)$ denotes the vitality at time t, n denotes the number of times that the node A have contacted other nodes. $Duration(i)$ denotes the time of the ith communication between node A and another node. $Interval(t)$ denotes an interval from the time that valid duration was recorded to the current time.

$$V_A(t) = \frac{\sum_{i=0}^{n} duration(i)}{interval(t)} \tag{6}$$

Consequently, the vitality and energy balance are utilized to obtain a "closer" and more reliable relay node in the proposed algorithm. In addition, differing from traditional routing strategy in terms of buffer management, old messages in buffer are deleted in the novel way instead of FIFO (First In First Out). In RCVE, when 99% of a node buffer space has been occupied, or the node wants to receive a new message which needs a bigger buffer space than its residual buffer space, the buffer management method will be called. Then the node will delete the old messages whose destination is neither in its local community nor the community of its contacting node.

4 Simulation Result

In this section, we will evaluate the performance of our RCVE based on the Opportunistic Network Environment (ONE) simulator [14]. We compare the RCVE with an improved probabilistic routing algorithm (PRTI) [8], Epidemic [7] and the routing based on simple community detection (RSCD) [9] in terms of four performance metrics, that is Average Delivery Ratio (ADR), Average Overhead (AO), Average Hop (AH), and Average Delay (AD). The simulation scenario is based on the built-in map in ONE, which is a real map of Helsinki downtown covering a 4500 m × 3400 m region. The map simulates and restores the streets, shops, parks, bus stops, and trams of Helsinki in detail. The nodes in the scenario are divided into six groups according to their different characteristics, such as moving speed, interface rate, buffer space, and so on. Specific parameter configurations are listed in Table 1. The message size varies from 500 KB to 1 MB. The duration of simulation is 12 h. Most importantly, for the fairness in contrast, the specific parameters such as p_i, α and β are reasonably set to get better results in PRTI, Epidemic, and RSCD.

Table 1. Parameters for nodes in different groups

Groups	Speed (m/s)	Pause time (s)	Communication range (m)	Data rate (kbps)	Buffer size (MB)
1,3	0.5, 1.5	0–120	10	250	10
2	2.7, 13.9	0–120	10	250	10
4	7, 10	10–30	1000	10000	50
5,6	7, 10	10–30	10	250	50

We evaluate the four algorithms under varying numbers of nodes from 5 to 35 for Group 1, Group 2, and Group 3, remaining the number of nodes in Group 4, Group 5, and Group 6 unchanged. The movement model is set as "ShortestPathMapBasedMovement" in which nodes get to their destination through the shortest valid path. The simulation results are shown in Fig. 2.

Figure 2(a) definitely shows the superiority of RCVE in terms of ADR followed by PRTI, RSCD, and Epidemic in sequence. Since the potential social attributes of nodes and the vitality are effectively utilized to improve the delivery rate. Meanwhile, the ADRs of other three methods nearly increase monotonically except Epidemic, as a result of its blindly forwarding messages.

Figure 2(b) depicts the average overhead of four algorithms, which has demonstrated that the RCVE is competitive inspect of AO on the basis of high delivery ratio. RCVE and RSCD deliberately select next nodes in communities to reduce redundant copies, which does not cause a great deal of overhead. However, Epidemic is based on flooding method, PRTI has the limitation of the number of replicas. Furthermore, RCVE with the energy balance in communities not only decreases the overhead in DTNs, but also avoids invalid options.

The AH of RCVE is outstanding among four algorithms as shown in Fig. 2(c). The main reason is that community division can improve the chances for nodes to find reliable relay nodes.

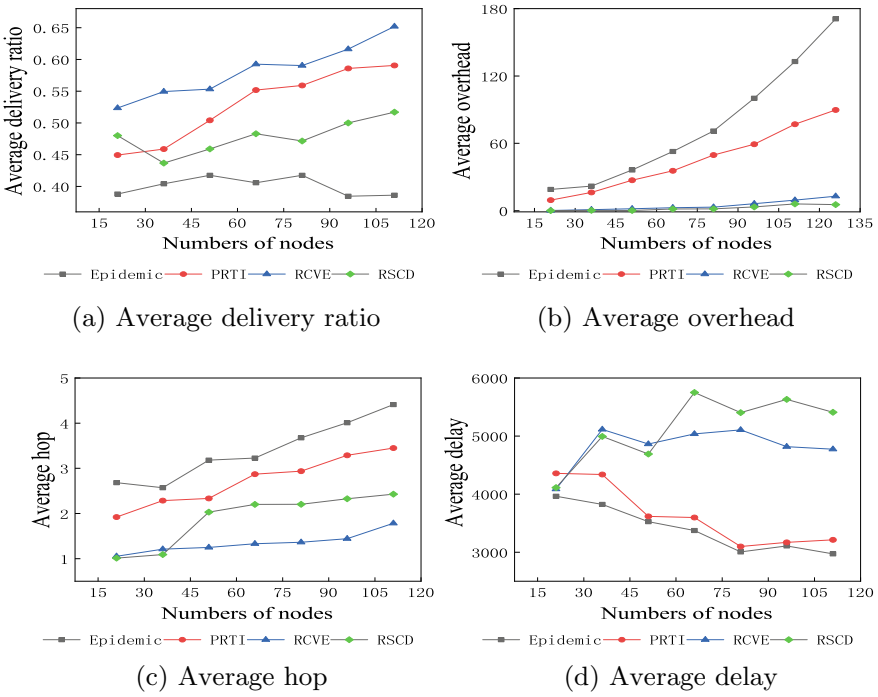


Fig. 2. Comparative results versus number of nodes.

Figure 2(d) compares the ADs of the four algorithms for different nodes. It is easy to see that the two algorithms based on community division have higher AD than others. The main reason is that RCVE and RSCD are devoted to find out the more suitable forwarders, and a longer delay may be incurred instead of blindly forwarding redundant copies to encountered nodes. Moreover, comparing with RSCD, AD of RCVE is reduced in virtue of effective buffer management and energy balance. Because in RCVE, suitable forwarders are able to forward more messages to their destinations during their lifetimes. In conclusion, the RCVE with efficient community division and forwarding method can improve the ADR with less overhead and hops.

5 Conclusion

In this paper, we proposed a novel routing protocol based on community division, vitality, and energy balance(RCVE), by utilizing the social attributes of nodes. In addition, controlling buffer and energy are simultaneously considered to select superior relay nodes. Finally, we compare RCVE with an improved probabilitic routing algorithm (PRTI), Epidemic and the routing based on simple community detection (RSCD) based on the ONE simulator. On the basis of experimental data, it is clear that RCVE can get superior forwarding path, and resists to buffer limitation, and has better performance in terms of ADR, AO, and AH.

References

1. Ezife, F., Li, W., Yang, S.: A survey of buffer management strategies in delay tolerant networks. In: IEEE International Conference on Mobile Ad Hoc and Sensor Systems, pp. 599–603 (2017)
2. Wang, H., Wang, S., Zhang, Y., Wang, X., Li, K., Jiang, T.: Measurement and analytics on social groups of device-to-device sharing in mobile social networks. In: 2017 IEEE International Conference on Communications (ICC), pp. 1–6. IEEE (2017)
3. Guo, T., Yang, Y.: Community based routing in social delay tolerant networks. In: 2015 Ninth International Conference on Frontier of Computer Science and Technology (FCST), pp. 321–324. IEEE (2015)
4. Garg, N., Rani, R.: A comparative study of community detection algorithms using graphs and R. In: 2017 International Conference on Computing, Communication and Automation (ICCCA), pp. 273–278. IEEE (2017)
5. Lu, Z., Sun, X., Wen, Y., Cao, G., La Porta, T.: Algorithms and applications for community detection in weighted networks. IEEE Trans. Parallel Distrib. Syst. **26**(11), 2916–2926 (2015)
6. Zhang, X., Wang, C., Su, Y., Pan, L., Zhang, H.F.: A fast overlapping community detection algorithm based on weak cliques for large-scale networks. IEEE Trans. Comput. Soc. Syst. (2017)
7. Zhang, X., Neglia, G., Kurose, J., Towsley, D.: Performance modeling of epidemic routing. Comput. Netw. **51**(10), 2867–2891 (2007)

8. Wang, X., He, R., Lin, B., Wang, Y.: Probabilistic routing based on two-hop information in delay/disruption tolerant networks. J. Electr. Comput. Eng. **2015**(3), 1–11 (2015)
9. Hui, P., Yoneki, E., Chan, S.Y., Crowcroft, J.: Distributed community detection in delay tolerant networks. In: MobiArch 2007: Proceedings of 2nd ACM/IEEE International Workshop on Mobility in the Evolving Internet Architecture ACM, New York, NY, USA, pp. 1–8 (2007)
10. Li, J., Wang, X., Cui, Y.: Uncovering the overlapping community structure of complex networks by maximal cliques. Phys. A Stat. Mech. Appl. **415**, 398–406 (2014)
11. Zhou, X., Liang, W., Wu, B., Lu, Z., Nishimura, S., Shinomiya, T., Jin, Q.: Dynamic community mining and tracking based on temporal social network analysis. In: 2016 IEEE International Conference on Computer and Information Technology (CIT), pp. 177–182. IEEE (2016)
12. Bulut, E., Szymanski, B.K.: Friendship based routing in delay tolerant mobile social networks. In: Global Telecommunications Conference, pp. 1–5. IEEE (2010)
13. Gregori, E., Lenzini, L., Mainardi, S.: Parallel k-clique community detection on large-scale networks. IEEE Trans. Parallel Distrib. Syst. **24**(8), 1651–1660 (2013)
14. Keranen, A., Andott, J., Karkkainen, T.: The one simulator for DTN protocol evaluation. In: International Conference on Simulation Tools and Techniques (Simutools 2009), pp. 55:1–55:10 (2009)

Performance Analysis of a Novel UAV Networks via Named Data Networking

Jiewen Cai[1(✉)], Zhuo Li[2], and Kangrong Wu[1]

[1] College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin 300387, China
jcai_tjnu@outlook.com

[2] School of Microelectronics, Tianjin University, Tianjin 300072, China
zli@tju.edu.cn

Abstract. Unmanned Aerial Vehicles (UAVs) are an emerging technology that can be utilized in military, public and civil applications. Multi-UAV systems can collaboratively complete missions more efficiently and economically as compared to single UAV systems. However, UAV networks in multi-UAV systems are still based on TCP/IP, which is not efficient and scalable to provide stable and reliable communication. Therefore, novel UAV networks via Named Data Networking (NDN) is proposed in this paper. Meanwhile, the simulation conducted on ndnSIM tests the transmission interference time and the maximum end-to-end transmission delay of this novel UAV networks, which are vital to the whole system. The result indicates the novel UAV networks via NDN satisfies the current network requirements, which has advantages such as good network adaptability, low latency, high security, etc.

Keywords: Named data networking (NDN) · Unmanned aerial vehicle (UAV) · Transmission interference time · End-to-end delay

1 Introduction

Unmanned aerial vehicles (UAVs), also commonly known as drones or remotely piloted aircrafts have found a wide range of applications during the past few decades [1]. With the enhancement of functions, UAVs play a more important role in many fields [2]. In some complex applications, UAVs are needed to work together, which can collaboratively complete missions more efficiently and economically compared to a single UAV [3]. Unlike many other wireless networks, the topology of UAV systems remains fluid with the number of nodes and links changing and also the relative positions of the nodes altering. The UAV networks would usually be deployed in dire cases and the network may get frequently partitioned, sometimes for long durations. Under these circumstances, traditional solutions do not guarantee connectivity.

Current IP network based on traditional communication system cannot meet the requirements of communications in terms of reliability, security, and mobility. Therefore, Van Jacobson of the Xerox Palo Alto Research Institute (PARC) proposed Named Data Networking (NDN) [4] for the deficiencies of the aforementioned IP network. Different from IP network, NDN is data-oriented and no longer contains

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 563–568, 2019.
https://doi.org/10.1007/978-981-13-6264-4_67

explicit location information such as a host or an interface address, which fundamentally changes the encapsulation structure and naming mechanism of IP packets, with short delay and low power consumption, high reliability, and other characteristics [5]. The advantages of NDN in content distribution, application-friendly communication, robust security, and mobility can be realized incrementally and relatively painlessly.

To tackle the existing problems in multi-UAV systems, this paper proposes here novel UAV networks via NDN to satisfy today's network requirement and introduces the characteristics of the system. In the Sect. 2, the background of NDN is described briefly. The architecture of this novel UAV networks via NDN is given in the Sect. 3. At last, this paper demonstrates the performance of UAV networks via NDN by switching connection modes among UAVs on ndnSIM [6, 7]. Simulation shows that transmission interference time is 431.2 ms and the maximum end-to-end transmission delay is 61.2 ms. The results indicate this novel UAV networks satisfy the current network requirements.

2 Named Data Networking

Communication in NDN is driven by the consumer. A consumer can retrieve an individual content object by sending a signaling message called Interest [8], which specifies the name of the desired content object. When a router receives an Interest, and has a copy of the content object in its local cache store (CS) [9], the router sends back the content without further propagating the Interest. If the router does not have the requested content, it looks up the next-hop neighbor(s) in the forwarding information base (FIB) [9] to forward the Interest to perform the longest prefix match of the name against its forwarding table. Whenever the Interest is relayed, an NDN router keeps the information in its pending interest table (PIT) [9] so that the response can be routed to the requester along the reverse path. Each entry in FIB maps a name prefix to a set of next-hop neighbor faces. The Interest is forwarded until it eventually reaches a node that has the requested content or it can generate it. This node sends back the content object which is propagated back to the consumer by following the reverse path of the Interest. Routers that forward a Data store it in their local caches so that they can directly respond to future requests.

3 Architecture of UAV Networks via NDN

The UAV networks belong to the distributed network and adopt the architecture of mobile Ad Hoc Network [10]. The entire architecture is generally divided into five layers: physical layer, link layer, network layer, transmission layer, and application layer. In order to achieve collaborative control, the data link of UAV is required to provide millisecond-level delay of information transmission and a higher transmission bandwidth. So, it is necessary for the architecture to adopt strong QoS guarantee. Due to the topology of multi-UAV system is multi-hop, non-centered and self-organizing structural, it is difficult to solve the delay guarantee problems of the UAV network layer.

The existing random competitive MAC protocols cannot shield against the conflict between the control packet and the data packet, and the delay cannot be effectively guaranteed. In addition, as the UAV networks belongs to the wireless Ad Hoc network, serious security problems will be caused by wireless channel, changing topology, etc.

In multi-UAV systems, the transmission layer mainly adopts TCP and UDP [11], and the network layer protocol is still the current research hotspot. Because NDN focuses on named data, adopts data-facing communication mode, and does not care where the contents are stored but the contents themselves, there is no need for IP address translation between the two systems. This new UAV networks architecture is based on the traditional framework of UAV systems, but NDN replaces TCP/IP in the network layer. Therefore, this section will mainly describe the process of this novel UAV networks at the network layer. When other UAVs request a certain content from a communication relay UAV, its packet communication process mainly includes the following four steps.

(a) The consumer encapsulates the data from application layer in Interest, and transmits the Interest to the data link layer and the physical layer, finally to the UAV's network.
(b) When the router in the network receives Interest, it further processes Interest according to the forwarding strategy of NDN.
(c) If Interest is not the same as the previous request, pass the Interest through UAV network forwards to the producer.
(d) The producer returns the Data corresponding to the Interest to the consumer.

4 Simulation

The frequent changes of topology are the major challenges faced by UAV communication. Therefore, this section will further demonstrate the performance of UAV networks via NDN by switching connection modes among UAVs on ndnSIM. As described in the Sect. 2, the transmitting interference time and the maximum end-to-end delay in the QoS service requirements are for the network layer performance indicators. Accordingly, the experiment will test these parameters of UAV networks based on NDN through specific simulation experiments to verify its performance.

4.1 Simulation Setting

The simulation is implemented in C++ and Python programming language, which tested on a PC with an Intel Core i5-3470 CPU@3.20 GHz and DDR3 SDRAM of 4 GB. In this experiment, the communication scene is simulated by setting relevant parameters on the platform ndnSIM, and the communication performance of this novel UAV networks is studied according to changes of UAV's topology. Among them, on the basis of the related indicators in the Link-16 data chain [12], the size of the packet transmitted in the network is set to 30 Byte, and the data transmission rate is 2 Mbps. The consumer sends 100 Interest each second to the producer. Routing and forwarding strategies of data are based on point-to-point module [13]. The experimental network topology is shown in Fig. 1.

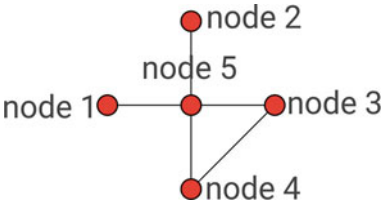


Fig. 1. Topology of simulation

In Fig. 1, node 1 represents the communication relay UAV and assumes the role of producer. Node 2, node 4, node 5 represent routers in NDN, and node 3 represents the consumer. In order to meet the actual communication conditions of the UAV, assume that the UAVs are running at a speed of 360 km/h. Within 0 to 2 s, node 3 communicates with node 1 through the router node 5, and disconnects with node 5 at 2.0000 s; and it is connected with node 4 at 2.3000 s. At this time, node 3 performs data transmission through node 4 and node 5 to node 1.

4.2 Simulation Results

4.2.1 Transmitting Interference Time

Before the communication link connects, the serial numbers changed linearly when node 3 received Data. It shows that NDN works normally. Node 3 received Data No. 195 at 1.9908 s and continued to send Interest to the producer in order from 1.9908 to 2.0000 s. When the producer receives the request, it returns the corresponding Data successively.

When node 3 connects with node 4 at 2.3000 s, node 3 starts to receive Data No. 224 at 2.3512 s. Later, the serial numbers increase linearly until Data No. 232 is received at 2.4312 s. Since Interest has a life cycle, only when the life cycle is over can the corresponding Interest in the PIT table be deleted. Therefore, the serial numbers of Data which is received from 2.4408 to 2.5412 s are fluctuant. It shows that the Interest from No. 196 to No. 204 have expired. The consumer can receive the corresponding Data only when he or she transmits these requests again. At the moment, the transmission process of communication link is disturbed. It goes back to normal until Data No. 223 is received at 2.7312 s. Therefore, in this experiment, NDN transmission interference time is 431.2 ms. The results can be seen in Fig. 2.

4.2.2 End-to-End Delay

End-to-end delay refers to how much time has elapsed since the packet left the source until it reached the end. When NDN stable transmission is 30 Bytes of Data, namely, a Data has only an Interest. When it fails to be requested again, maximum end-to-end transmission delay of the system is less than 61.2 ms. The results are shown in Fig. 3. The end-to-end delay was 40.8 ms before the communication link breaks. When it connected again, the end-to-end delay becomes 61.2 ms, which caused by longer route path.

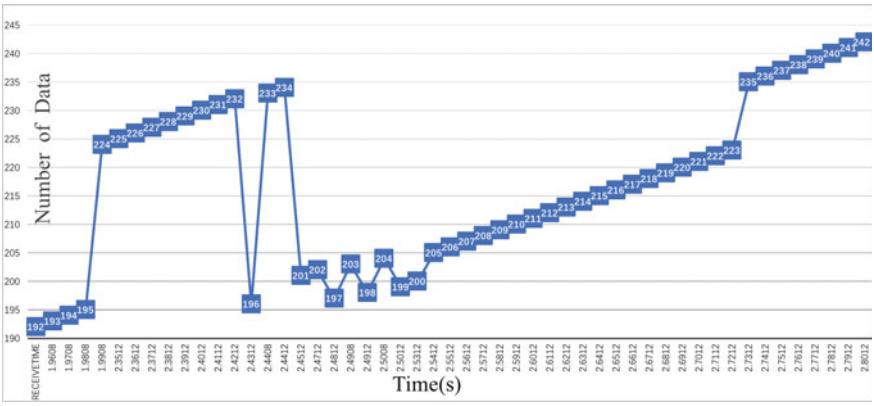


Fig. 2. Transmitting interference time

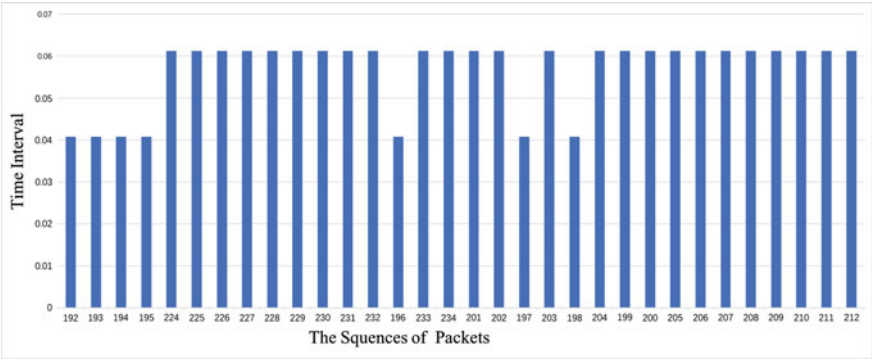


Fig. 3. End-to-end delay

The simulation result shows that this UAV networks via NDN has good communication performance in mobile communication. Its transmission interference time is less than 61.2 ms. The caching function of NDN improves data transmission rate and network resources utilization as well as communication reliability. At the same time, NDN transmits data based on content name, which greatly improves communication mobility. Even when the disconnection occurs, the communication can be resumed quickly. The communication losses which are caused by a link failure can be reduced to minimum.

5 Conclusion

This paper proposes a novel UAV networks via NDN, and tests the transmitting interference time and the end-to-end delay, which are fatal to efficient and reliable communication. The simulation results show that the UAV networks via NDN has excellent performance, and can fully cope with the future data communication needs of multi-UAV systems.

Acknowledgement. This work was supported by the National Students for Innovation and Entrepreneurship Training Program of China under Grant (201810065011).

References

1. Valavanis, K.P., Vachtsevanos, G.J.: Future of unmanned aviation. Handbook of Unmanned Aerial Vehicles. Springer Netherlands, pp. 2993–3009 (2015)
2. Zeng, Y., Zhang, R., Lim, T.J.: Wireless communications with unmanned aerial vehicles: opportunities and challenges. IEEE Commun. Mag. **54**(5), 36–42 (2016)
3. Gupta, L., Jain, R., Vaszkun, G.: Survey of important issues in UAV communication networks. IEEE Commun. Surv. Tutor. **18**(2), 1123–1152 (2016)
4. Zhang, L., et al.: Named data networking (ndn) project. Relatório Técnico NDN-0001, Xerox Palo Alto Research Center-PARC (2010)
5. Mosko, M., Solis, I., Wood, C.A.: Content-Centric Networking-Architectural Overview and Protocol Description (2017). arXiv:1706.07165
6. Afanasyev, A., Moiseenko, I., Zhang, L.: ndnSIM: NDN simulator for NS-3. University of California, Los Angeles, Technical Report 4 (2012)
7. Mastorakis, S., et al.: ndnSIM 2.0: a new version of the NDN simulator for NS-3. NDN, Technical Report NDN-0028 (2015)
8. Saxena, D., et al.: Named data networking: a survey. *Comput. Sci. Rev.* **19**, 15–55 (2016)
9. Saxena, D., et al.: Reliable memory efficient name forwarding in named data networking. In: 2016 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering (DCABES). IEEE (2016)
10. Bouachir, O., et al.: A mobility model for UAV ad hoc network. In: 2014 International Conference on Unmanned Aircraft Systems (ICUAS). IEEE (2014)
11. Kung, H.T., et al.: A location-dependent runs-and-gaps model for predicting TCP performance over a UAV wireless channel. In: Military Communications Conference, 2010-MILCOM 2010. IEEE (2010)
12. Howitt, S., Platts, J.: Real-time deep strike mission simulation using air-launched UAVs. In: 1st UAV Conference (2002)
13. Zhu, Z., Afanasyev, A., Zhang, L.: A new perspective on mobility support. Named-Data Networking Project, Technical Report (2013)

Construction of Distributed Virtual Backbone Network in Tactical Communication Network

Zhikai Zhang[1,2] and Fusheng Dai[1,2(✉)]

[1] Harbin Institute of Technology at Weihai, Weihai 264209, China
dfs7113@126.com
[2] Science and Technology on Communication Networks Laboratory, Shijiazhuang 050081, China

Abstract. With the modernization of warfare and the development of electronic reconnaissance and confrontation capabilities, communication nodes or links in the network topology of tactical communication are becoming more and more vulnerable to the enemy's precise attack. When the network nods are destroyed, the corresponding link is no longer present. At the same time, the network's business needs and flow dredging needs will be increasing with the worsening of the war. So, it will bring delays to the network, resulting in the quality of networks services declining continuously. In order to solve such problems, this paper adopts the method of constructing virtual backbone network to realize. The important information, higher priority information and high real-time information (such as voice, video, etc.) are dredged through a virtual backbone network, and other information is dredged through a common physical network. Virtual backbone network is constructed by using the existing physical network topology and the resources reserved by the physical network, and can connect directly in the network layer. Virtual backbone network can be understood as a special channel for important information and high real-time information.

Keywords: Tactical communication network · Virtual backbone network · Network topology construction

1 Introduction

The construction technology of virtual backbone network is a key issue in the research and discussion of experts and scholars at home and abroad in recent years. The main algorithm is divided into two major categories: Construction of virtual backbone network for centralized computing and construction of virtual backbone network for distributed computing. The two algorithms have their own advantages and disadvantages. The advantage of centralized algorithms is that the whole virtual network can be constructed through central node or command node, but the demand for the central node is high, and the whole information of global network needs to be controlled in real time, so it brings the communication overhead. The advantage of the distributed algorithm is that it can divide the computing pressure of the central node into all the nodes in the network, Each node does not need to know the global information only need to know the perimeter node location and resource information, reducing the

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 569–576, 2019.
https://doi.org/10.1007/978-981-13-6264-4_68

computational pressure and communication overhead. The disadvantage is that the computational algorithm is more complex and the computing nodes are too dispersed.

According to the characteristics of tactical communication network and the complexity and unpredictability of battlefield environment, this paper adopts distributed algorithm. Because the battlefield environment is unpredictable, the network topology will change, and the distributed computing method can improve the computational speed. At the same time, distributed computing avoids the failure of a central node in centralized computing to cause a network crash.

The construction of the traditional virtual backbone network is realized by constructing the connected dominating set, the idea is to find the maximal independent set (MIS) of the network first, then construct the connected dominating set (CDS) according to the maximal independent set. The connected dominating set as the virtual backbone network. In addition, this paper also puts forward the calculation idea, also is to obtain the maximal independent set first, then constructs the virtual network according to the disjoint path algorithm, finally mapping between a virtual network and a physical network, obtains the final virtual network. In the paper [1], the search strategy is used to find the connected dominating set, and the results are cropped, the algorithm is simple but the computation is too large. The paper [2] proposes a distributed algorithm based on cluster, which does not solve the problem of literature [1] fundamentally. In the paper [4–6], the concept of structural fault tolerance is introduced into the backbone network, which can effectively guarantee the node's undamaged. The paper [7, 8] proposes to increase the node to establish the virtual backbone network with connectivity conditions. In the paper [9], a virtual backbone network with connectivity condition is constructed by using the method of Steiner tree instead of adding nodes on the basis of the literature [7, 8].

2 Construction of Virtual Backbone Network Based on Distributed Algorithm

2.1 Mathematical Model of Backbone Network

N nodes are randomly generated in the two-dimensional plane, and the maximum transmission radius of each node can be different, because of the full-duplex characteristic of the tactical communication network, the arbitrary link of the generated graph is bidirectional, that is, the generated network model is the undirected graph, the network model is shown in Fig. 1. In Fig. 1, the maximum transmission radii for a, b, and c three points are r_a, r_b and r_c respectively. The dotted circle represents the maximum transmission radius of the respective nodes, when the distance between node b and node a is less than the maximum transmission radius of r_a, it is considered that node a and b can communicate directly, that is, the existence of link. When the distance between node c and node a is greater than the maximum transmission radius of r_a, the node c and node a cannot communicate directly, that is, there is no link.

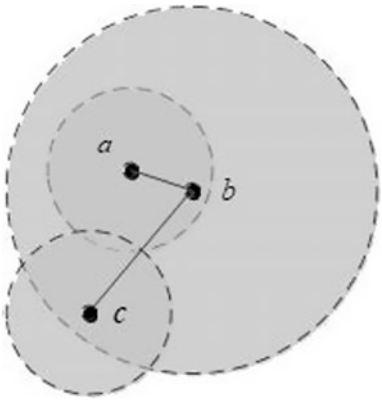


Fig. 1. Network node model

2.2 The Construction Process of Virtual Backbone Network

According to the foregoing, constructing the maximal independent set is the first step of constructing the virtual backbone network. First of all, we should search the neighboring node information, in the microwave network topology, the method of acquiring the neighboring nodes by any node is as follows: all nodes in the network topology broadcast "hello" information to neighboring nodes; after a round of broadcast, any node in the whole network topology can know the information of the neighboring nodes, and the adjacent information table can be obtained; repeat the above operation

$$\begin{cases} k(p) > k(q) \\ k(p) = k(q), tab(p) > tab(q) \end{cases} \Rightarrow W(p) > W(q) \tag{1}$$

$$\begin{cases} k(p) < k(q) \\ k(p) = k(q), tab(p) < tab(q) \end{cases} \Rightarrow W(p) < W(q) \tag{2}$$

R (v) indicates whether the adjacent node is a leaf node, and R (V) = 0 indicates that its adjacent node is not a leaf node, and R (V) = 1 indicates that its neighboring node is a leaf node.

The construction of maximal independent set can be understood as the process of dividing all nodes in network topology into dominating nodes and controlled nodes.

$$T(v) = \begin{cases} 0 \Rightarrow \text{The current node is not yet sorted} \\ 1 \Rightarrow \text{The current node is the dominant node} \\ 2 \Rightarrow \text{The current node is the dominated node} \end{cases} \tag{3}$$

Introducing parameter G(v), G'(v), and H(v), G(v) represent the set of adjacent nodes of node v, G'(v) represents the set of R = 1 in the adjacent nodes of node v, and H(v) represents the set of T = 0 in the adjacent node of node v.

The steps to build a maximum set of independent methods are as follows:

Step 1: For any network node v, to determine whether the node received the dominant message, if you receive the control message, the T(v) set to 1, otherwise proceed to step 2.

Step 2: Determine if R(v) is 0 by judging the adjacent nodes of v, and if R(v) equals 0, go to step 4, or proceed to step 3.
Step 3: Continue to judge G(v), if G(v) is an empty set, sets T(v) to 2, and then broadcasts the dominant message, if G(v) is not an empty set, enter step 5.
Step 4: Continue to judge G'(v), if G'(v) is the empty set, go to step 6, or else T(v) to 2, and then broadcast its dominant message.
Step 5: Traverse G(v) to have a node with a weight greater than v, and if so, keep the state of the node unchanged, and if not, set T(v) to 2 and broadcast its dominant message.
Step 6: Traverse H(v), if there is a node with a weight greater than v, if the T (v) is not set to 2, and then broadcast its dominant message, otherwise, end the algorithm.

All nodes in the microwave network topology use the above steps, the set of all t = 2 points is a maximal independent set [3], and the maximal independent set can be obtained by the above method.

Step 1: Introduce the parameter Y and make the following settings.

$$Y(v) = \begin{cases} 0 \Rightarrow \text{The current node is not yet sorted} \\ 1 \Rightarrow \text{The current node joins the CDS} \end{cases} \tag{4}$$

At the same time, each control node sends invitations, the invitation information includes the following: The node's tag $tab(v)$ and the timer counter, the initial value is set to 3.
Step 2: Determine if the node that receives the information is the dominant node, and if it is the dominant node, go to step 3, or go to step 4.
Step 3: The Y of the current node is set to 1, and the "return" message is sent to the original control node, which is sent back along the path of the invitation information and the reverse path is set to 1 for all nodes in the control node path.
Step 4: Log the $tab(v)$ tag of the current microwave network node to the invitation information, and then continue forwarding.

Take each of the control nodes above, and loop through the Steps 1–4, until all the Y = 1 points are composed of connected dominating sets, ending the algorithm.

3 Simulation Analysis of Network Construction

(1) Topological structure output

Assuming the combat range of tactical communication network is 100 km × 100 km, the number of microwave stations of tactical communication network is 20, for the convenience of calculation, the maximum transmission radius of microwave station is set to 60 km. Figure 2 gives the topology structure of network based on disjoint path, and obtains the topological structure of Fig. 3 as MIS by constructing maximal independent set algorithm. The blue solid node is the selected dominating point, and the network topology of CDS is formed through the connection of the dominating point. As shown in Fig. 4, the blue solid node and the corresponding bold link form the topological structure of the backbone network.

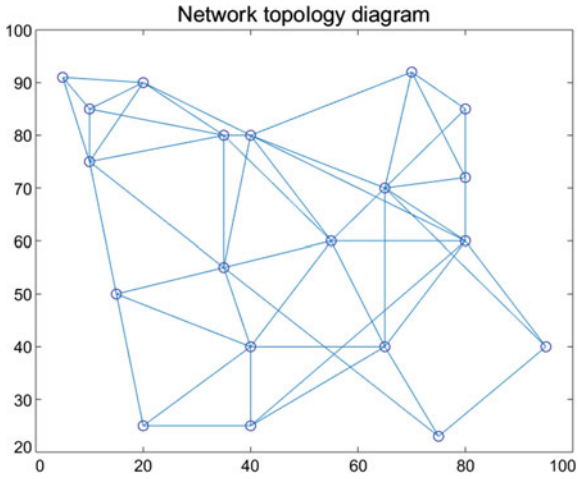


Fig. 2. Stochastic topological graph generated by disjoint paths

(2) Performance analysis

In order to highlight the performance of the algorithm, we select the traditional backbone network construction algorithm WAN algorithm [3] for comparison, set the number of iterations 100 times, Fig. 5 for this algorithm and the traditional algorithm WAN algorithm on the number of CDS relative nodes of the comparison.

As can be seen from Fig. 5, when the number of nodes is the same, the algorithm in this paper is less than the number of relative nodes obtained by the other two traditional methods, which shows that the backbone network is smaller in size. At the same time, the relative node number of the two algorithms decreases with the increase of the number of nodes. It is indicated that when the number of nodes increases, there are

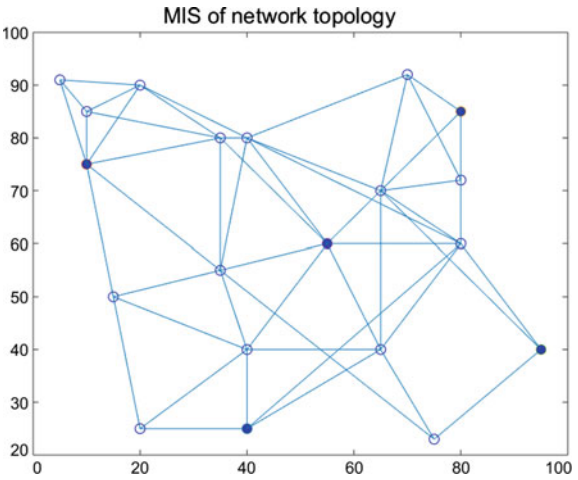


Fig. 3. The topological structure of MIS

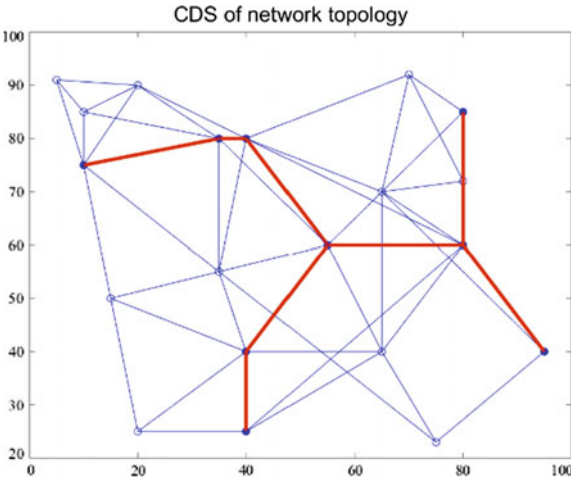


Fig. 4. The topological structure of CDS

more nodes in the network, and the connectivity between nodes increases, so that the number of nodes does not need to be the dominating point and the connected dominating set is formed, so the relative nodes decrease with the increase of the total number of nodes. It is concluded that the construction algorithm of backbone networks mentioned in this paper is superior to that of traditional backbone networks.

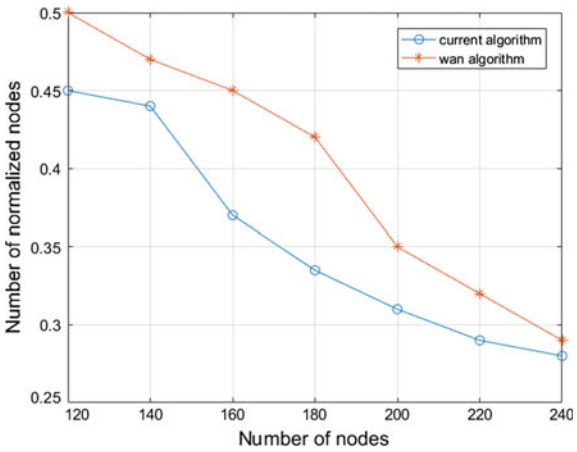


Fig. 5. The topological structure of CDS

4 Conclusion

In order to guarantee the deterioration of service quality caused by the network damage under the war condition, in this paper, a method of constructing virtual backbone network under the condition of microwave networks is proposed, which is to compute the maximal independent set first, and then to solve the connected dominating set to construct the virtual network. By comparing with the traditional construction method, it is found that under the precondition of satisfying the same business level, this algorithm can effectively build a smaller backbone network, so it proves the performance and feasibility of the algorithm.

References

1. Wu, J.: Extended dominating set based routing in ad hoc wireless networks with unidirectional links. IEEE Trans. Parallel Distrib. Syst. **13**(9), 866–881 (2002)
2. Wu, J., Li, H.: A dominating set based routing scheme in ad hoc wireless networks. Telecomm Syst. Spec. Issue Wirel. Netw. **18**(1–3), 13–36 (2001)
3. Wan, P.J., Alzoubi, K.M., Frieder, O.: Distributed construction of connected dominating set in wireless ad hoc networks. In: Joint Conference of the IEEE Computer & Communications Societies, pp. 141–149 (2004)
4. Basu, P., Redi, J.: Movement control algorithms for realization of fault tolerant ad hoc robot networks. IEEE Netw. **18**(4), 36–44 (2004)
5. Li, N., Hou, J.C.: A fault-tolerant topology control algorithm for wireless networks. In Proceedings of International Conference on Mobile Computing and Networking, pp. 275–286 (2004)
6. Li, X.Y., Wan, P.J., Wang, Y., Yi, C.W.: Fault tolerant deployment and topology control in wireless networks. In: Proceedings of Mobile Ad Hoc Networking & Computing, pp. 117–128 (2003)

7. Bredin, J.L., Demaine, E.D., Hajiaghayi, M., Rus, D.: Deploying sensor networks with guaranteed capacity and fault tolerance. In Proceeding of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing, pp. 309–319 (2005)
8. Koskinen, H., Karvo, J., Apilo, O.: On improving connectivity of static ad hoc networks by adding nodes. Mediter. Ad Hoc Netw. **197**(3), 169–178
9. Agrawal, A., Klein, P., Ravi, R.: When trees collide: an approximation algorithm for the generalized steiner problem on networks. Technical Report CS-90-32, Brown University (1991)

Multipath TCP Scheduling Optimization Based on PSRBP in Heterogeneous Network

Haitao Zhao(✉), Mengkang Zhang, Hongsu Yu, Tianqi Mao, and Hongbo Zhu

College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China
zhaoht@njupt.edu.cn

Abstract. Nowadays, with the large-scale deployment of multiple network interfaces in heterogeneous networks, the diversity of interfaces makes the characteristic differences between transmission paths tend to be obvious. However, when a heterogeneous network uses MPTCP for transmission, a large amount of out-of-order data packets accumulate in the receive buffer due to the different characteristics of each path, causing congestion of the receive buffer and resulting in poor MPTCP transmission performance. Aiming at the problem of MPTCP receive buffer blocking, this paper proposes a multipath scheduling optimization scheme named Path Selection and Receive Buffer Prediction (PSRBP)). The scheme predicts throughput based on the throughput prediction model, introduces parameter that indicates the similarity of other paths to the path with the highest throughput, and compares it with the established threshold to select the paths that can be transmitted by MPTCP. Afterward, the proposed scheme estimates the size of the receive buffer by judging the number of out-of-order packets and predicts the buffer blocking in advance and stops the data transmission on the bad paths to improve the performance of MPTCP. The simulation results show that the proposed scheme can improve the overall throughput of the system, making the network utilization rate improved.

Keywords: Heterogeneous network · MPTCP · Receive buffer blocking · Multipath scheduling optimization

1 Introduction

With the advancement of science and technology, the network tends to be heterogeneous. Collaboration and cooperation among multiple networks largely satisfy the communication requirements and user service diversity requirements in the mobile environment [1]. The MultiPath Transmission Control Protocol (MPTCP) is a protocol proposed by the Internet Engineering Group (IETF) in 2009 [2]. It is an extension protocol of the Transmission Control Protocol (TCP), allowing the simultaneous use of multiple paths for data transmission. Using MPTCP can improve resource utilization. MPTCP uses TCP as a subflow, with each path representing each TCP subflow [3]. MPTCP can achieve higher throughput by increasing the efficiency of network resource

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 577–589, 2019.
https://doi.org/10.1007/978-981-13-6264-4_69

usage. Applying MPTCP to a heterogeneous network can improve transmission throughput.

One of the major challenges in using MPTCP is the receiver buffer blocking problem [4]. The data scheduling algorithm adopted by MPTCP default is the Round-Robin algorithm. The algorithm sends data to each subflow in turn by polling, without considering the differences between subflows [5]. However, because of the differences of delay, bandwidth, load, packet loss rate, and other factors between each path in heterogeneous network, using the Round-Robin algorithm simply will cause packets to be out of order. Out-of-order packets accumulate in a limited receive buffer, which can cause blocking problem and severely affect the communication performance between mobile terminals. In this case, it is very important to design an effective multipath scheduling optimization method to avoid the receiver buffer blocking problem and improve the transmission performance of the heterogeneous network.

2 MPTCP Receive Buffer Blocking

The MPTCP receive buffer blocking problem is a research hotspot in recent years. A large number of scholars and research institutions attempt to reduce the receiving buffer by means of data scheduling, packet reordering, link scheduling, and path management, which can improve system throughput.

The authors of [3] propose an MA-MPTCP mobile-aware multimedia data transmission algorithm using MPTCP, which can effectively improve the stability and fairness of using MPTCP transmission, and slow down the receiving buffer blocking problem. The authors of [6] study the blocking problem of the MPTCP receive buffer, and propose a scheme to avoid redundant retransmission, which can prevent the receive buffer from being blocked and improve the throughput of the overall system. However, this scheme requires the establishment of multiple connections for each interface in advance, resulting in the waste of resources. In [7], there is a new data scheduling strategy named MPTCP-MA^2 for the MPTCP receiver buffer blocking problem, which can use the path more properly, but the effect is not obvious. The authors of [8] propose a route scheduling on asymmetric path connections, which can reduce the out-of-ordered packets in the receive buffer and improve the robustness of the system. However, they only deal with single-flow transmission experiment, but the effect on multi-flow transmission is still unknown. A new PR-R algorithm is proposed in [9] to reorder packet in MPTCP. This algorithm can improve the total throughput and path utilization of the system in the case of high packet loss rate. However, it occupies a large memory space and is not convenient enough for practical applications. The authors of [10] combine the packet reordering recovery method with the congestion control algorithm to improve the path utilization, but the throughput improvement for the asymmetric link is not obvious. The authors of [11] compare different flow scheduling strategies. The results show that scheduling data flow to a specific path is better than time-slice rotation scheduling. However, their work is based on multi-flows, and there is limited guidance for the transmission of single-flow in a path of greater variability.

Aiming at the problems occurred in the solution of solving the buffer blocking, we propose an MPTCP multipath scheduling optimization scheme based on path selection

and buffer prediction in heterogeneous networks. The proposed scheme can avoid the buffer blocking problem by predicting the buffer blocking phenomenon and stopping the transmission of poor performance paths. It can improve the system throughput and maximize the performance of MPTCP.

3 Path Scheduling Optimization Based on Path Selection and Buffer Prediction

3.1 Path Selection Based on Throughput Prediction and Available Bandwidth

Different wireless communication technologies exist in heterogeneous networks. Different terminals also have different network interfaces. When the heterogeneous network uses multipath transmission, there are great differences of characteristics between these paths. MPTCP creates a TCP subflow for each path. These subflows share the same receive buffer, and their different path characteristics affect MPTCP performance. In order to prevent buffer blocking problem and achieve higher throughput, more data packets should be transferred by the paths with the best characteristics. If the performance of a path is poor, the throughput increment caused by the aggregate bandwidth is smaller than the throughput reduction caused by its RTT and packet loss rate, so this path should be discarded. This section proposes a path selection method based on throughput prediction and available bandwidth to compare the performance of each path and stop data transmission through poor performance subflows.

The conventional throughput model based on RTT and packet loss rate is

$$T = \frac{1}{RTT} * \sqrt{\frac{3}{2bp}} * packet\_size \tag{1}$$

where RTT is the round trip time of the path, p is the packet loss rate, b is the number of packets acknowledged by the received ACK, where $b = 1$ is assumed. This formula can predict the throughput of each path under the standard TCP protocol.

Let $I = \{1,2,...,n\}$ represent all available paths. The parameters of these paths are $path_i = \{RTTi, P_i, BW_i\}$. Among them, RTT_i is the round trip time of available path i, P_i is the packet loss rate of available path i, and BW_i is the available bandwidth of available path i. The maximum throughput is

$$T_{\max} = \max_{i \in I} T_i \tag{2}$$

In the case where the available bandwidth of the best path is limited, there are

$$T_k = T_{\max} \text{ and } T_k > BW_k \tag{3}$$

where T_k is the throughput of path k, BW_k is the available bandwidth of path k, the throughput of path k is equal to the maximum throughput, and its throughput is greater than its available bandwidth. At this point, the set $S_0 = \{k | T_k = T_{\max} \text{ and } T_k > BW_k\}$ is the initial path selection scheme.

Assume that BW_{km} is the maximum available bandwidth in S_0, X_{max} is the throughput of the best path, path i does not belong to the initial path selection scheme, X_i represents the throughput of path i, and has the following expression:

$$BW_{km} = \max_{k \in S_{init}} BW_k \tag{4}$$

$$X_{\max} = T_{\max} = T_k \tag{5}$$

$$X_i = T_i (i \notin S_0) \tag{6}$$

As shown in Eq. (7), when $j \notin S_0$ and $j \neq 0$, $\Delta_{m,j}$ represents the difference between T_{max} and the throughput of other paths. When $j = 0$, $\Delta_{m,j}$ represents the difference between T_{max} and the maximum bandwidth of each optimum path.

$$\Delta_{m,j} = X_{\max} - X_j \tag{7}$$

Let $\delta_{m,j}$ represent the similarity between X_{max} and X_j, then there are

$$\Delta_{\max} = \max \Delta_{m,j}, \Delta_{\min} = \min \Delta_{m,j} \tag{8}$$

$$\delta_{m,j} = \frac{\Delta_{\min} + \zeta \Delta_{\max}}{\Delta_{m,j} + \zeta \Delta_{\max}} \tag{9}$$

The bigger $\delta_{m,j}$ is, the closer X_{max} is to X_j. Among them, $\zeta \in [0, 1]$ is a distinction factor. When $\Delta_{\max}$ becomes too large, ζ will weaken the influence of $\Delta_{\max}$. Here, let ζ = 0.5.

δ_{thr} is a threshold that determines whether a path is selected. A path will be selected when $\delta_{m,i} \geq \delta_{thr}$. When X_0 is large, fewer paths are needed to aggregate the bandwidth. When X_0 is smaller, more paths are needed to aggregate the bandwidth. Therefore, here is:

$$\delta_{thr} = \delta_{m,0} \tag{10}$$

When X_0 is large, δ_{thr} will also be larger, only fewer paths will be selected at this time. Conversely, when X_0 is larger, more paths will be selected.

In summary, the overall path selection scheme is

$$S = S_0 \cup S_1 \tag{11}$$

$$S_1 = \left\{ i | \delta_{m,i} \geq \delta_{thr} \right\} \tag{12}$$

According to the path selection method based on throughput prediction and available bandwidth, we can select paths with higher similarity to the best path and stop the transmission of packets through the paths with poor performance.

3.2 PSRBP Path Scheduling Optimization Scheme

According to the path selection method based on throughput prediction and available bandwidth proposed in the previous section, we manage the available paths and propose a path scheduling optimization scheme based on path selection and receive buffer prediction (PSRBP), which can effectively prevent receive buffer blocking and increase system throughput. We refer to MPTCP using this path management scheme as PSRBP-MPTCP.

If the receive buffer size is large enough, buffer blocking will not occur, so there is no difference between PSRBP-MPTCP and conventional MPTCP. To enable PSRBP-MPTCP to operate like conventional MPTCP, the required receive buffer size is

$$Buffer = 2 * \sum_{i=1}^{N} BW_i * RTT_{\max} \tag{13}$$

Where *Buffer* represents the required receive buffer size, N is the number of subflows, *BWi* represents the available bandwidth of the ith subflow, RTT_{max} refers to the largest *RTT* in all subflows. Because each path has different characteristics, this receive buffer size is not practical. In a heterogeneous network, each destination does not have enough receive buffer size for MPTCP. The conventional MPTCP always tries to use all available paths, ignoring the difference of characteristics of the path and the actual available buffer size, and therefore suffers from the problem of receive buffer blocking. In the absence of enough receive buffers and different characteristics of multiple paths, in order to ensure system throughput, path management scheme based on path selection and buffer prediction should be used.

First, predict receive buffer blocking in advance. Receive buffer blocking occurs when the buffer size required by MPTCP exceeds the available receive buffer size. To predict buffer blocking, the number of out-of-order packets and the required buffer size must be estimated [5]. During the RTT of subflow j, the number of out-of-order packets estimated in subflow i is

$$\begin{aligned} L_{i,j} &= Floor\left(\frac{RTT_j}{RTT_i}\right) - 1, \frac{RTT_j}{RTT_i} \geq 2 \\ L_{i,j} &= 0, \frac{RTT_j}{RTT_i} < 2 \end{aligned} \tag{14}$$

At this point, $j > i$. For fast estimation, assume that the number of packets that have been transmitted during the *RTT* is one transmission unit. Under this assumption, $L_{i,j}$ represents the expected number of transmission units associated with out-of-order packets in subflow i during RTT_j. To avoid excessive transmission constraints, let $L_{i,j} = 0$ when the number of out-of-ordered packets during RTT_j is less than the number of packets transmitted during RTT_i. Because the proposed algorithm estimates the number of out-of-order packets transmitted during each packet transmission, the number of untransmitted packets during the *RTT* is equal to the subflow congestion window

(*cwnd*). Based on $L_{i,j}$, the buffer size required for MPTCP to transmit using all available paths at the same time is

$$B_{\text{request}} = \sum_{R(N)} PA_i * L_{i,N} * MSS_i \tag{15}$$

$$R(N) = \{i \in Z | 1 \leq i \leq N\} \tag{16}$$

Where PA_i is the untransmitted data packets in subflow i, MSS_i is the largest message segment in subflow i, N is the number of subflows, and subflow i represents the subflow being used. According to the congestion control algorithm, *cwnd* can be changed during the RTT of each subsflow. However, because it is difficult to predict changes in *cwnd*, we assume that the *cwnd* of all subflows is fixed during the RTT of the subflows.

Once the required buffer size is estimated, the PSRBP algorithm can predict buffer blocking. If buffer blocking is predicted, it will find that subflows with poor performance are found according to the path selection algorithm proposed in the previous section and stop transmitting through these subflows. Due to the real-time changes in the topology of the heterogeneous network, it is possible to stop some of the available paths by mistake. If the available receive buffer size is large enough, all the stopped subflows are used again to send the packets, i.e., when the available buffer size of the destination exceeds 2.5 times the required buffer size, we define that the available buffer size is large enough. Algorithm 1 shows the overall algorithm of PSRBP.

Algorithm 1 Operation of PSRBP

```
for all subflows j do
  if cwnd of best performance subflow is not filled
    transfer through best performance subflow
  else transfer through subflows in the S0 collection
recalualte:
    calculate B_request
    if B_request > available buffer then
      calculate δ_{m,j}
       if δ_{m,i} ≤ δ_thr
         make path j inactive
         goto recalculate
       end if
    else
      if available buffer > 2.5*B_request
        use all the subflows
      else
      use subflows in the S_0 collection
      end if
    end if
  end if
end for
```

4 Simulation Results and Analysis

This section evaluates the performance of PSRBP-MPTCP and conventional MPTCP in heterogeneous networks. This experiment implements PSRBP-MPTCP in the multipath TCP-Linux kernel version 0.89. In order to generate a virtual heterogeneous network environment, we use NS-3 for network simulation. This simulation is equipped with four interfaces supporting MPTCP, which can transmit packets from the sending end to the receiving end. The simulation path topology is shown in Fig. 1. There are four disjoint paths between two MPTCP terminals, represented as path A, path B, path C, and path D, respectively. The access bandwidth of path A is 1 Mbps, and Path A has a 100 ms transmission delay, and has a packet loss rate of 0.02. The access bandwidth of path B is 10 Mbps, path B has a 50 ms transmission delay, and has a packet loss rate of 0.01. The access bandwidth of path C is 3 Mbps, path C has a 40 ms transmission delay, and has a packet loss rate of 0.01. Path D has an access bandwidth of 2 Mbps, has a transmission delay of 20 ms, and the packet loss rate is 0.01.

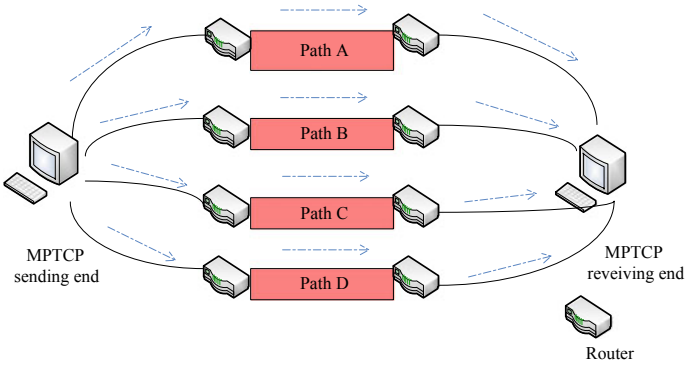


Fig. 1. Simulation path topology

The MPTCP receive buffer size is set to 200 KB, the path management method is set to fullmesh, the congestion control algorithm is reno, and opportunistic retransmission and penalty mechanism are disabled. The main parameter settings are shown in Table 1. The source and destination are connected through four disjoint paths, and each path is connected through two routers. The link of source-to-router and the link of router-to-destination have sufficient performance and do not affect the overall throughput (bandwidth: 100 Mbps, delay: 1 ns). The queue type and size of each link are Drop-Tail and 100 packets, and the simulation time is 100 ms. The main simulation parameters of the path are shown in Table 1.

Table 1. Main path simulation parameters

Parameter	Path A	Path B	Path C	Path D
Bandwidth	1 Mbps	10 Mbps	3 Mbps	2 Mbps
Propagation delay	100 ms	50 ms	40 ms	20 ms
Queue limit	100	100	100	100
Queue type	Drop-Tail	Drop-Tail	Drop-Tail	Drop-Tail
Path management	Fullmesh	Fullmesh	Fullmesh	Fullmesh
Congestion control	reno	reno	reno	reno
Packet loss rate	0.002	0.001	0.001	0.002

Figure 2 shows the throughput of four-path PSRBP-MPTCP using paths A, B, C, and D. As shown in the Fig. 2, the bandwidth of the subflow B is fully utilized, and the remaining three subflows are subject to different degrees of restriction. This is because the path selection algorithm considers the subflow B to have the best performance. Since using all the subflows to transmit data packets will cause the receiver buffer blocking, the PSRBP-MPTCP will not allocate the data packets to the subflow A and subflow C where the performance is poor for transmission. When the data packets transmitted through the subflow D increase, a blocking phenomenon of the receive buffer occurs, and at this time, the subflow D is restricted from being transmitted, and all data packets are transmitted through the subflow B. When only one subflow is

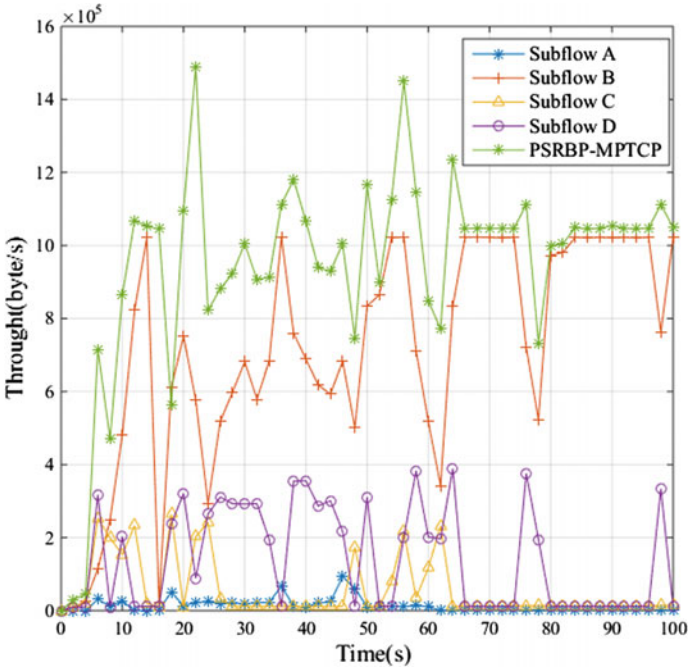


Fig. 2. Throughput of PSRBP-MPTCP using path A, B, C, and D

transmitted, the receive buffer blocking phenomenon will disappear. Once the available receiving buffer size is greater than 2.5 times the prediction buffer size, all subflows are reused for transmission. The total throughput of PSRBP-MPTCP is greater than the throughput using single-flow transmission, and the performance does not deteriorate.

Figure 3 shows the throughput of four-path conventional MPTCP using paths A, B, C, and D. When using the same characteristic paths as PSRBP-MPTCP, the performance of conventional MPTCP is significantly worse than that of PSRBP-MPTCP. Since the conventional MPTCP does not include our proposed PSRBP algorithm, all subflows are used for transmission at the beginning, and the total throughput reaches the maximum at 10 s. With the continuous transmission of data packets, the receive buffer gradually fills up and blocking occurs. Each subflow is limited to sending data packets, and buffer blocking continue to occur, so the throughput has been at a low level, losing the significance of multipath transmission.

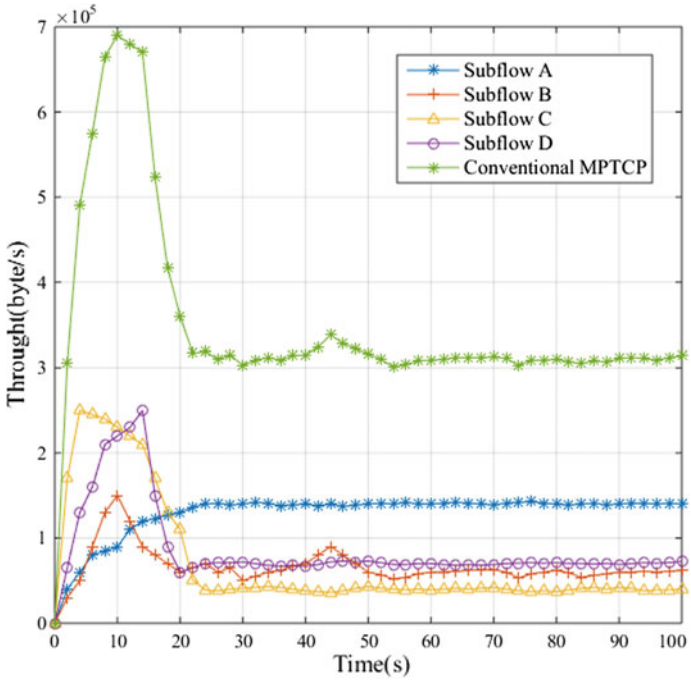


Fig. 3. Throughput of conventional MPTCP using path A, B, C, and D

Figures 4 and 5, respectively, use the accumulated curves to obtain the throughput of PSRBP-MPTCP and conventional MPTCP using paths A, B, C, and D, respectively. As can be seen from Fig. 5, the accumulated throughput of both PSRBP-MPTCP and conventional MPTCP increases linearly within 100 s, but the growth rate of PSRBP-MPTCP is faster, and the accumulated throughput achieved at 100 s is also higher than that of conventional MPTCP, which means network utilization has improved.

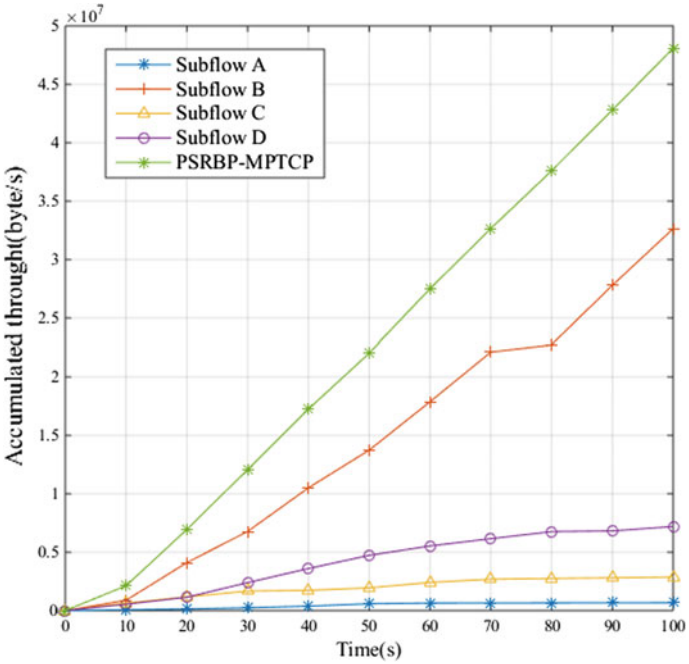


Fig. 4. Accumulated throughput of PSRBP-MPTCP using path A, B, C, and D

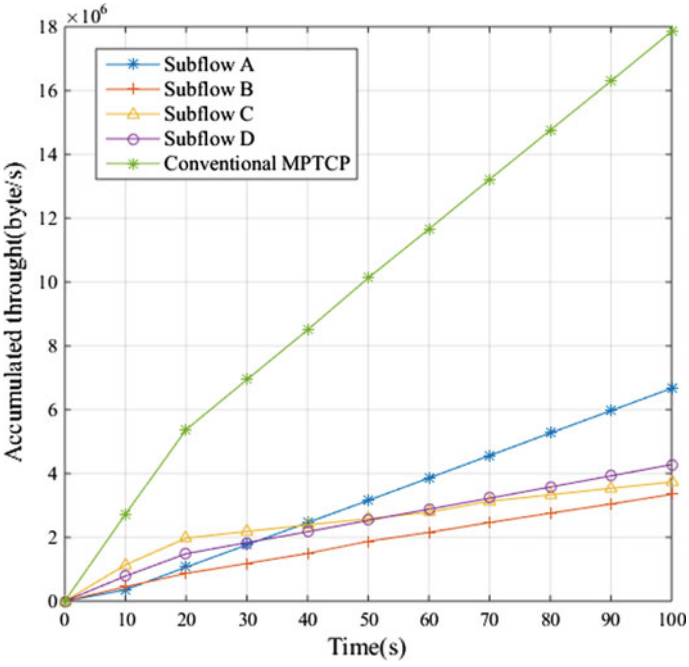


Fig. 5. Accumulated throughput of conventional using path A, B, C, and D

To illustrate the performance of PSRBP-MPTCP, this paper compares the PSRBP-MPTCP with other improved algorithms. Figures 6 and 7 compare the throughput and accumulated throughput of PSRBP-MPTCP with MA-MPTCP, MPTCP-MA^2, and conventional MPTCP, respectively. As shown in Fig. 6, in the first 30 s, RSRBP-MPTCP does not perform well enough, and the advantage is not obvious compared to MA-MPTCP and MPTCP-MA^2, and even the performance is worse than the other two algorithms. After 30 s, PSRBP-MPTCP tends to be stable, and its throughput is significantly greater than that of other algorithms. As shown in Fig. 7, in the 0–50 s phase, PSRBP-MPTCP has little difference with the accumulated throughput of other algorithms. In the 50 s–100 s phase, PSRBP-MPTCP has significantly improved accumulated throughput compared to the other two improved algorithms.

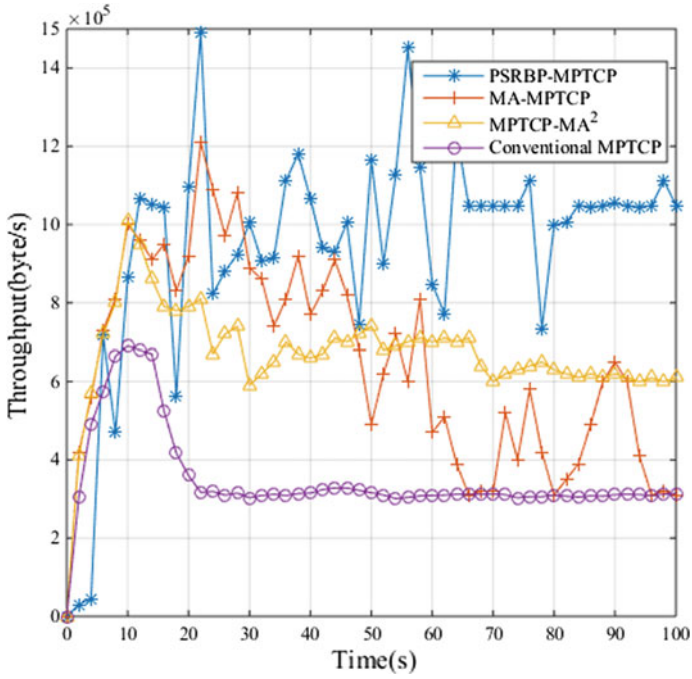


Fig. 6. Throughput comparison of PSRBP-MPTCP, MA-MPTCP, MPTCP-MA^2, and conventional MPTCP

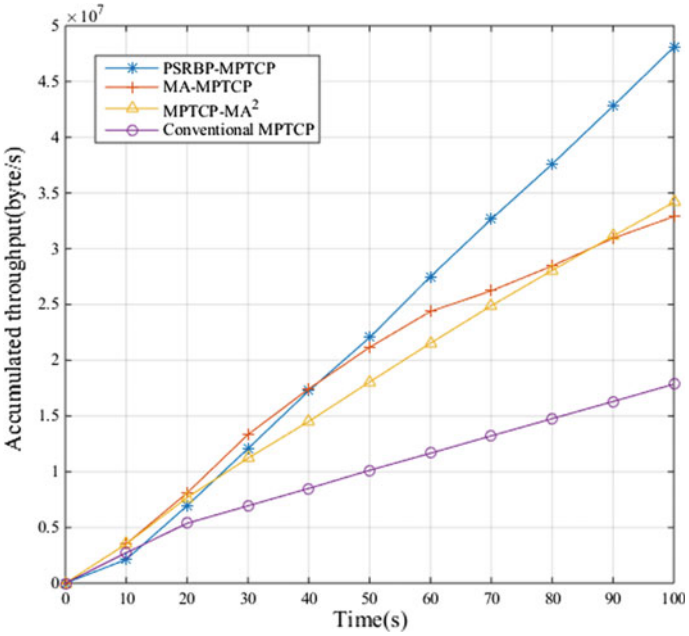


Fig. 7. Accumulated throughput comparison of PSRBP-MPTCP, MA-MPTCP, MPTCP-MA^2, and conventional MPTCP

5 Conclusion

Aiming at the problem of MPTCP receive buffer blocking, this paper proposes a multipath TCP path scheduling optimization scheme based on path selection and buffer prediction to improve the throughput of MPTCP and network utilization in heterogeneous networks. The proposed scheme estimates the number of out-of-order packets, thereby determining the size of the required receive buffer and comparing it with the actual available buffer size, so that it can predict the receive buffer blocking. Once buffer blocking is predicted, the path selection algorithm based on the throughput model and available bandwidth is used to stop transmission over poor performance paths. By this method, the receiver buffer blocking problem will be reduced, and PSRBP-MPTCP will have better performance than conventional MPTCP, achieving an overall increase in throughput and improving network utilization.

Acknowledgment. This work was supported by National Natural Science Foundation of China (61771252, 61471203), Basic Research Program of Jiangsu Province (BK20171444), "The Six Talents High Peaks" Funding Project of Jiangsu Province (DZXX-041), "1311" Talents Funding Project of Nanjing University of Posts and Telecommunications, the Practice Innovation Program for College Graduates of Jiangsu Province (No.JSCX17_0224).

References

1. Karam, F.W., Jensen, T.: Performance, analysis of ranking for QoS handover algorithm for selection of access network in heterogeneous wireless networks. In: International Conference on Computer Communications and Networks, pp. 1–6. IEEE (2012)
2. Zhu, D., Xu, C., Qin, J., et al.: Mobility-aware multimedia data transfer using multipath TCP in vehicular network. In: Wireless Communications and Mobile Computing Conference, pp. 1041–1046. IEEE (2017)
3. Becke, M., Adhari, H., Rathgeb, E.P., et al.: Comparison of multipath TCP and CMT-SCTP based on intercontinental measurements. In: Global Communications Conference, pp. 1360–1366. IEEE (2014)
4. Oh, B.H., Lee, J.: Constraint-Based Proactive Scheduling for MPTCP in Wireless Networks. Elsevier North-Holland Inc. (2015)
5. Barré, S.: Implementation and Assessment of Modern Host-Based Multipath Solutions. Bonaventure Olivier (2011)
6. Özcan, Y., Guillemin, F., Houzé, P., et al.: Fast and smooth data delivery using MPTCP by avoiding redundant retransmissions. In: IEEE International Conference on Communications, pp. 1–7. IEEE (2017)
7. Ke, F., Huang, M., Liu, Z., et al.: Multi-attribute aware multipath data scheduling strategy for efficient MPTCP-based data delivery. In: Communications. IEEE (2016)
8. Sysomphone, C., Pattarawit, P., Chatchai, K.: The path scheduling for MPTCP end-to-end hosts. In: International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, pp. 127–130. IEEE (2017)
9. Alheid, A., Doufexi, A., Kaleshi, D.: Packet reordering response for MPTCP under wireless heterogeneous environment. In: International Conference on Telecommunications, pp. 1–6. IEEE (2016)
10. Alheid, A., Doufexi, A., Kaleshi, D.: A study on MPTCP for tolerating packet reordering and path heterogeneity in wireless networks. In: Wireless Days, pp. 1–7. IEEE (2016)
11. Dreibholz, T., Seggelmann, R., Tüxen, M., et al.: Transmission scheduling optimizations for concurrent multipath transfer. In: International Workshop on Protocols for Future, Large-Scale and Diverse Network Transports, pp. 2074–5168 (2010)

Study on Connected Target Coverage Algorithm for Wireless Sensor Network

Danyang Qin(✉), Jingya Ma, Yan Zhang, Pan Feng, and Ping Ji

Key Laboratory of Electronics Engineering, College of Heilongjiang University, Harbin 150080, Heilongjiang, China
qindanyang@hlju.edu.cn,
{majingya659, fengpan_1993}@163.com,
zhangyanheida@126.com, 1539892132@qq.com

Abstract. Aiming at the severe effect of coverage hole induced by the limited energy and the harsh deployment environment of wireless sensor network (WSN) on network lifetime, an optimized energy-efficient-connected target coverage heuristic (OECCH) algorithm with lightweight is presented. Based on the analysis of contribution to coverage, the sensor nodes are classified to achieve different functions, then an intersection cover set is constructed based on heuristic values, so as to maximize the network lifetime. Simulation results indicate that OECCH can prolong network lifetime while reducing computation overhead of WSN obviously.

Keywords: Wireless sensor network · Connected target coverage · Coverage/connectivity heuristic · Network lifetime

1 Introduction

Recent breakthroughs in wireless communication technology and digital electronics have greatly spurred the emergence of WSN, and WSN has been widely applied to a variety of range of applications [1]. With being deployed in a harsh environment, the battery-powered sensor nodes are intractable to be charged or replaced, and node failures have easily occurred because of energy depletion [2]. Topology control is one of the most basic problems in WSN. It is of great significance for prolonging network lifetime [3]. And the connected target coverage (CTC) problem has become an important part in the design of topology control. This thesis mainly studies how to maximize network lifetime on the basis of CTC problem [4].

There are many classical researches on how to maximize network lifetime by scheduling sensor nodes. Cardei et al. proposed a connected set covers (CSC) algorithm [5], this algorithm used impractical energy consumption model which is not fit for the actual network environment. An approximate algorithm named CWGC is proposed by Zhao and Gurusamy [6], the algorithm has low dead nodes ratio. However, there is no solution to the target point with the lower coverage rate. The OCCH algorithm proposed by Zorbas and Douligeris [7] is the closest to our proposed algorithm. This algorithm prolongs the network lifetime considering the connectivity issue. However, this algorithm is only suitable for densely deployed network. A maximum-connected load-balancing coverage tree (MCLCT) is discussed in [8], but what it adopted is a disjoint cover set, in which each

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 590–597, 2019.
https://doi.org/10.1007/978-981-13-6264-4_70

sensor can only generate one cover set. While, in the intersection cover set, each sensor node is able to join multiple cover sets. In other words, the intersection set can form more cover sets than that of the disjoint set. In this paper, an intersection cover set is formulated according to heuristic value for prolonging the lifetime of WSN. Considering the energy consumption in the network and overcoming the above shortcomings, an optimized energy-efficient-connected target coverage heuristic algorithm (OECCH) is carried out.

2 Problem Description

2.1 Maximum Cover Tree (MCT) Problem

Provided that there are N sensor nodes distributed around M number of target points randomly. Let $S = \{s_1, s_2, \cdots, s_N\}(|S| = N)$ and $P = \{p_1, p_2, \cdots, p_M\}(|P| = M)$ denote a set of sensor nodes and a set of target points (TPs), respectively. Each TP in set P is supposed to be covered by at least one sensor node so as to maximize the network lifetime, and the initial energy e_0 of all sensor nodes in S is the same. A two-dimensional undirected weighted graph $G = (V, E)$ is constructed for the given WSN to describe the topology of the initial network, where V is composed of S, P and a set of sink R, i.e.,$V = S \cup R \cup P, E = \left\{ (s_i, s_j) \middle| d_{i,j} \le r_c \right\}$ denotes the set of edges. s_i and s_j will be said to be connected on the edge of undirected weighted graph G if the Euclidean distance $d_{i,j}$ between any two nodes (say s_i and s_j) is no greater than communication range r_c.

Let $T(\tau) = (S_s(\tau) \cup S_r(\tau), E(\tau))$ express a constructed tree in time interval τ, where $S_s(\tau)$ and $S_r(\tau)$ represent a set of sensing nodes and a set of relay nodes respectively, and $S_s(\tau) \cap S_r(\tau) = \varnothing$, $E(\tau)$ represents an edge set used for connecting active nodes S_a and the sink R. Sensing nodes in $S_s(\tau)$ perform monitoring task, while relay nodes in $S_r(\tau)$ are used for relaying the sensed data to the sink R. It can be seen from Fig. 1 that the constructed cover tree has three characteristics: (1) the root of the tree is sink R; (2) each leaf of the tree is a sensing node; (3) each target point is covered by at least one sensing node. In conclusion, the energy consumption model for each node in $T(\tau)$ can be classified as follows:

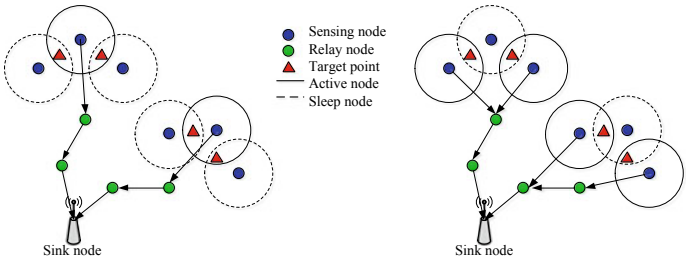


Fig. 1. Examples of MCT

$$E(s, T(\tau)) = \begin{cases} e_{sens} + e_{trans} B(\tau), & s \in S_s(\tau), s \notin S_r(\tau) \\ (e_{trans} + e_{rec}) B(\tau) D(s, T(\tau)), & s \in S_r(\tau), s \notin S_s(\tau) \end{cases} \quad (1)$$

where $T(\tau)$ represents the constructed cover tree in time interval τ, $B(\tau)$ is the number of data packets collected by nodes in time interval τ, $D(s, T(\tau))$ denotes the number of descendant nodes of node s in $T(\tau)$, e_{sens} and e_{rec} express the energy consumption for

perceiving and transmitting 1-bit data, respectively, e_{trans} indicates the energy consumption for transmitting 1-bit data from the sender s_i to receiver s_j.

Definition 1 (MCT problem): Given a undirected-connected graph $G = (V, E)$, and the initial energy for each node in the graph is e_0, where $V = S \cup P \cup R$, and E is an edge set. Our objective is to maximize the network lifetime LT through finding a family of cover tree $T(\tau_1), T(\tau_2), \ldots, T(\tau_w)$ and its corresponding time interval $\tau_1, \tau_2, \ldots, \tau_w$. Thus, the MCT problem can be mathematicalized as follows, where $\lambda_{x,y}$ is an observation variable which will be defined in Sect. 2.2.

$$\textbf{Maximization}: \text{Max } LT \equiv \sum_{i=1}^{w} \tau_i$$

$$\textbf{Subject to}: \sum_{i=1}^{w} E(s, T(\tau_i)) \leq e_0, \forall s \in S$$

$$\sum_{x \in S_s(\tau_i)} \lambda_{x,y} \geq 1, \ \forall y \in P, \tau_i \in [1, \tau]$$

2.2 Coverage and Connectivity Model

If the distance between arbitrary node $s_i \in S$ and the target point $p_j \in P$ is no greater than the sensing radius r_s of node s_i, p_j is called to be covered by s_i, noted as $\lambda_{i,j}$, where the distance between any two nodes can be deduced from the Euclidean distance formula. Thus, we have

$$\lambda_{i,j} = \begin{cases} 1, & \|s_i - s_j\| \leq r_s \\ 0, & \|s_i - s_j\| > r_s \end{cases} \tag{2}$$

To avoid cover redundancy at the TPs, the coverage ratio should be verified. The coverage ratio CR of s_i is the number of TPs covered by s_i divided by the total number of TPs, that is

$$CR(s_i) = \frac{\left| \bigcup_{s_i \in S} F(s_i) \right|}{|P|} \tag{3}$$

If the distance between sensor s_l and s_m $(s_l, s_m \in S)$ is less than the communication range r_c, s_l and s_m will be termed as connected. Let $\varphi_{l,m}$ denote the connectivity between s_l and s_m, and $\varphi_{i,m}$ can be defined as

$$\varphi_{l,m} = \begin{cases} 1, & \|s_l - s_m\| \leq r_c \\ 0, & \|s_l - s_m\| > r_c \end{cases} \tag{4}$$

3 An Optimized Energy-Efficient-Connected Target Coverage Heuristic (OECCH) Algorithm

An optimized energy-efficient-connected target coverage heuristic (OECCH) algorithm will be presented in this section. Our main objective is to find maximum number of intersection cover set (ICS), alternately activating the cover sets in ICS, the network lifetime is maximized

along with the adoption of energy update. OECCH algorithm will find the cover set C_i as much as possible until the nodes in S_a (set of active nodes) cannot form any new independent cover sets. Thus, S_a is divided into several subsets by the cover sets generated by OECCH algorithm $(C_i \cap C_j \neq \{\varnothing\}, i \neq j)$. The nodes in each cover set C_i can collaboratively monitor all TPs. Hence, the collection of active nodes subsets can be obtained, i.e., $ICS = \{C_1, C_2, \ldots, C_w\}$. The specific process of OECCH algorithm is shown is Algorithm 1.

Algorithm 1 Connected Coverage optimization heuristic algorithm

Input: S_s, S_r, a family of $F(s)$
Output: *ICS*, *LT*
(1)**Begin**
(2) $ICS = \{\varnothing\}$; $L = \{\varnothing\}$; LT=0; i=1;
(3) **repeat**
(4) **while** $F_u(s) \neq P$
(5) $presct\_F_u = \{\varnothing\}$, $presct\_node = \{\varnothing\}$, $C_i = \{\varnothing\}$;
(6) **for** each $s_i \in S_s$ **do**
(7) calculates $CvH(s_i)$ according to Eq.(5);
(8) $presct\_node = \max_{s_i}(CvH)$;
(9) **if** $|CR(C_i \cup presct\_node)| > |CR(presct\_F_u)|$
(10) $presct\_F_u = F_u(C_i \cup presct\_node)$;
(11) $C_i = C_i \cup presct\_node$;
(12) **end if**
(13) **end for**
(14) **for** each $s_k \in S_r$ **do**
(15) **if** $\|s_k - R\| > r_c$ **then**
(16) $S_r = S_r \setminus s_k$;
(17) $ncr = ncr \cup s_k$;
(18) $one\_hop = S_r$;
(19) **end if**
(20) **for** each $s_k \in one\_hop$ **do**
(21) calculates $CnH(s_k)$ according to Eq.(6);
(22) $presct\_node = \max_{s_k}(CnH)$;
(23) $C_i = C_i \cup presct\_node$;
(24) **end for**
(25) **end while**
(26) **until** $F_u(C_i) = P$
(27) **for** each $s_i \in C_i$
(28) $L(s_i) = e_i / e_{pc}$;
(29) $L = L \cup L(s_i)$;
(30) **end for**
(31) $ICS = ICS \cup C_i$;
(32) $L(C_i) = \min(L)$;
(33) $LT = LT + L(C_i) \times t_{stamp}$;
(34) $i = i + 1$
(35) **return** *ICS*
(36) **return** *LT*
(37)**end**

Let $F(s_i)$ represent the coverage of s_i for TPs $(i = 1, 2, \ldots, n)$, and let $F_u(s)$ denote the union of several $F(s_k)$, where $k = 1, 2, \ldots, q$ $(q \leq n)$. We hope that $F_u(s)$ can cover all the TPs, i.e., $F_u(s) = F(s_1) \cup F(s_2) \cdots \cup F(S_q) = P$. First, the coverage heuristic value (CvH) of a sensing node $(s_i \in S_s)$ will be calculated according to its contribution to the coverage, as is shown in Eq. (5). Then, the sensing node with maximum CvH will be regarded as the preselected node, which can be expressed as $presct\_node = \max_{s_i}(CVH)$. If the coverage ratio has been improved with the participation of the preselected node, updates the preselected coverage $(presct\_F_u)$ and places the preselected node into C_i (line 4–13).

$$CvH(s_i) = \sum_{j=1}^{|P|} \frac{\lambda_{i,j}}{\sum_{k=1}^{|S|} \lambda_{k,j}} \tag{5}$$

$$CnH(s_k) = \frac{\sum_{i=1}^{|ncs|} \frac{\varphi_{m,i}}{\sum_{n=1}^{|S_r|} \varphi_{n,i}}}{\|s_k - R\|} \tag{6}$$

Furthermore, we calculate the connectivity heuristic value (CnH) of the relay nodes that are connected to the sink directly, as is shown in Eq. (6), where ncs denotes the set of sensor nodes that are not connected to the sink R, ncr expresses the set of relay nodes that are not connected to sink R. The relay node with higher CnH will be the preselected node and will be placed into C_i (line 14–24). Repeat the above steps until the network is fully covered, an intersection cover set $ICS = \{C_1, C_2, \ldots, C_w\}$ can be eventually obtained.

Let e_i be the total energy consumption of the sensor node, and e_{pc} express the average energy consumption per lifecycle. L is the set of lifecycle of nodes in arbitrarily cover set C_i. It can be observed that the minimum value in set L is the lifecycle of C_i, i.e., $L(C_i) = \min(L)$. Thus, the network lifetime (LT) of the WSN can be derived (line 27–36). Furthermore, the energy of entire selected sensors is updated, i.e., for $\forall s_i \in C_i$, if $e_i > e_{pc}$, s_i will perform sensing or relaying task; otherwise, the energy of node s_i is depleted, and s_i will be removed from C_i.

4 Simulation Results and Performance Analysis

The proposed OECCH algorithm is simulated and analyzed based on MATLAB platform compared with the existing connected coverage algorithms, which are MCLCT, OCCH, CWGC, and CSC, respectively. For rigorous observation, the reference connected coverage algorithms adopt the same energy consumption model with the proposed OECCH algorithm. The simulation analysis mainly focuses on the maximization of the network lifetime in comparing the OECCH algorithm.

4.1 Performance Analysis of Network Lifetime

The effects of number of sensors (NS), number of target points (NT) and communication ability of sensors (CAS) on network lifetime will be analyzed in this section. Provided that the size of interested area is 150 m × 150 m, the position of the sink node

is at (0, 75). The communication radius of any sensor node is $r_c = 50\,\text{m}$, and the sensing radius of any sensing node is $r_s = 10\,\text{m}$. Each simulation scenario is executed 50 times with a confidence interval of 95%. Provided that all sensor nodes have the same initial energy $e_0 = 20\,\text{J}$. The energy required by the sensing node in data sensing is the same as that of data receiving by the relay node, that is $e_{sens} = e_{rec} = 100\,\text{nJ/bit}$. What is more, the system energy consumption is $e_t = 50\,\text{nJ/bit}$, and the energy consumption per-hop is $e_{hop} = 100\,\text{pJ/(bit} \cdot \text{m)}$.

Network lifetime with different NS is plotted in Fig. 2, where the number of sensor nodes changes from 150 to 600 with an increment of 50. *y*-axis represents the network lifetime and the unit of it is hour (h). Simulating results in Fig. 2 illustrate that the network lifetime extends with the increasing of sensor nodes. In addition, it can be observed that the network lifetime of the proposed OECCH algorithm is the longest when the number of sensor nodes is fixed.

The effect of NT on the network lifetime is plotted in Fig. 3. There are 300 sensor nodes being deployed around the TPs, where the number of TPs varies from 5 to 50 with an increment of 5. As we can see from Fig. 3, the network lifetime reduces with the increase in the number of TPs for all five algorithms, and the CSC algorithm has the worst performance. The network lifetime of the proposed OECCH algorithm is the longest when the number of TPs is fixed.

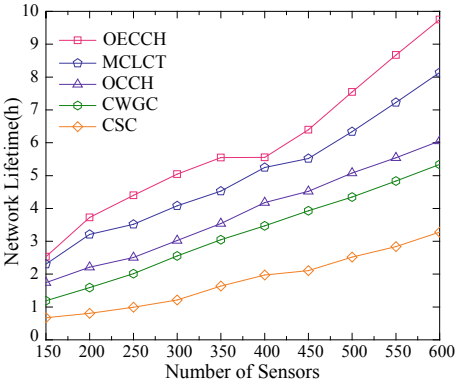


Fig. 2. Network lifetime with different NS

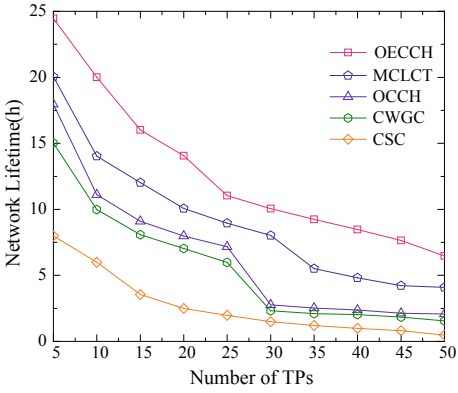


Fig. 3. Network lifetime with different NT

In order to analyze the impact of CAS on the network lifetime, we assume that there are 30 TPs in the network, and the communication radius of a sensor node varies from 10 m to 110 m with an increment of 30 m. As is shown in Fig. 4, the network lifetime extends as the communication radius increases. This owing to the increase of the communication radius of the sensor node, which improves the connectivity of the network and reduces the use of the relay nodes, and the lifetime of WSN extends as a consequence of the increase in cover set. As we can see from Fig. 4, the OECCH algorithm is superior to the existing MCLCT, OCCH, CWGC, and CSC algorithms.

In conclusion, the OECCH algorithm will prolong the network lifetime efficiently compared with the other four algorithms.

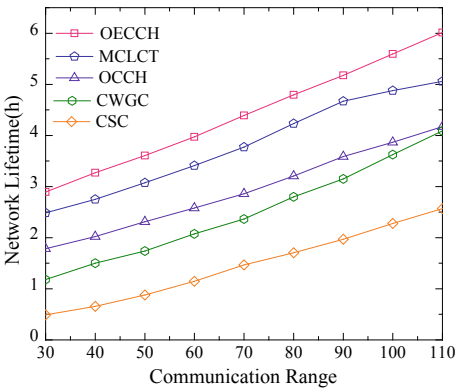


Fig. 4. Effect of CAS on network lifetime

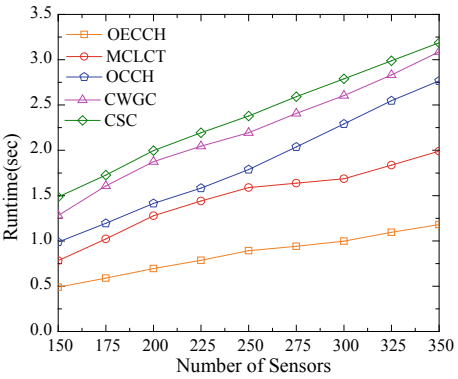


Fig. 5. Runtime with different NS

4.2 Runtime

Runtime is one of the most important indicators of computation overhead. The runtime of the OECCH algorithm and the existing algorithm under with different numbers of sensor nodes is given in Fig. 5, where the unit of runtime is second (sec), and the number of sensor nodes vary from 150 to 350. It can be observed from Fig. 5 that the OECCH algorithm runs faster than the other four algorithms. In the OECCH algorithm, the sensing node is only responsible for sensing information, while the relay node is only responsible for forwarding information, which will save a lot of energy and reduce the runtime. In addition, compared with the existing algorithms, the number of cover sets is being increased due to the adaptation of intersection cover set, and the cover set selection is relatively simple and direct, which avoids the coverage redundancy at TPs.

In summary, compared to the existing MCLCT, OCCH, CWGC, and CSC algorithms that are widely used in WSN, the proposed OECCH algorithm can prolong the lifetime of WSN efficiently and reduce the computation overhead obviously.

5 Conclusions

In order to extend the lifetime of WSN, an OECCH algorithm is presented in this paper. In WSN, nodes are formed into the largest number of subsets to monitor all target points and the subset will be activated one by one for monitoring targets. The proposed OECCH algorithm follows the QoS metric. The intersection cover set is built based on the coverage or connectivity of sensor nodes. The proposed algorithm takes precautions against the coverage redundancy at the TPs. We divide the sensor nodes into relay nodes and sensing nodes to avoid unnecessary traffic and protect the energy from early exhaustion. The proposed scheme avoids the premature emergence of coverage and

connectivity vulnerabilities in the WSN, thus expanding the network lifetime. The network lifetime of OECCH algorithm is 166–172% longer than that of CSC algorithm, 42–46% longer than that of CWGC algorithm, 14–36% longer than that of OCCH algorithm, and 10–15% longer than that of MCLCT algorithm. It is obvious that the proposed OECCH algorithm is superior to the existing connected coverage algorithm.

Acknowledgement. This work was supported in part by National Natural Science Foundation of China (61771186), Postdoctoral Research Project of Heilongjiang Province (LBH-Q15121), University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province (UNPYSCT-2017125).

References

1. Sanoob, A.H., Roselin, J., Latha, P.: Smartphone enabled intelligent surveillance system. IEEE Sens. J. **16**(5), 1361–1367 (2016)
2. Qin, D., Yang, S., Zhang, Y., et al.: Heuristic algorithm of lifetime maximization for wireless sensor network. In: International Conference in Communications, Signal Processing, and Systems, pp. 341–351 (2016)
3. Deif, D.S., Gadallah, Y.: Classification of wireless sensor networks deployment techniques. IEEE Commun. Surv. Tutor. **16**(2), 834–855 (2014)
4. Kim, Y.H., Han, Y.H., Jeong, Y.S., et al.: Lifetime maximization considering target coverage and connectivity in directional image/video sensor networks. J. Supercomput. **65**(1), 365–382 (2013)
5. Cardei, I., Cardei, M.: Energy-Efficient Connected-Coverage in Wireless Sensor Networks. Inderscience Publishers, pp. 201–210 (2008)
6. Zhao, Q., Gurusamy, M.: Lifetime maximization for connected target coverage in wireless sensor networks. IEEE/ACM Trans. Netw. **16**(6), 1378–1391 (2008)
7. Zorbas, D., Douligeris, C.: Connected coverage in WSNs based on critical targets. Comput. Netw. **55**(6), 1412–1425 (2011)
8. Chen, C., Mukhopadhyay, S.C., Chuang, C., et al.: Efficient coverage and connectivity preservation with load balance for wireless sensor networks. Sens. J. IEEE **15**(1), 48–62 (2015)

Protocol Design of Efficient Access for Laptop in Wi-Fi Networks with Guaranteed Security

Juan Zhao(✉) and Xiaolong Xu

Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing 210003, PR China
juanzhao.njupt@gmail.com, xuxl@njupt.edu.cn

Abstract. Computer security is very important since three decades ago, which becomes quite challenging especially in the wireless access scenarios. In this paper, we summarize the technologies of the development over the computer security. Then, we propose a couple of iterative algorithms to carry out the laptop selection to wirelessly access the computer networks with the existence of the eavesdropper. The goal of designing the algorithms is, respectively, to maximize the secured throughput of the wireless access channel and maximize the energy efficiency for the target laptop with the existence of the eavesdropper. Simulations demonstrate the effectiveness of the proposed algorithms.

Keywords: Computer networks · Guaranteed safety · Energy efficient access · Wi-Fi networks

1 Introduction

The security issue has been the key problem in computer networks since three decades ago. In [1], the researcher from Naval Research Laboratory summarizes the existing methods by 1983 for the computer security, where both of the military and the commercial computer networks are taken into the consideration. Both the rules and the actions are addressed in [1] to construct a secured computer network. In [2], the definition of computer security are clarified. The strategies is discussed. The professional techniques and the administrative mechanism are explained. These results are from the IBM research center located in Los Angeles. In [3], the network protocol is emphasized for computer network security, where the development of the corresponding protocol is still an open problem by the year 1987. In [4], the definition of computer security is updated in the year 2014, and the insecurity factors are comprehensively analyzed. Both of the attack techniques and the protection techniques are summarized. The

This paper is partially supported by the Scientific and Technological Support Project (Society) of Jiangsu Province under Grant BE2016776, the National Key Research and Development Program of China under Grant 2018YFB1003702.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 598–604, 2019.
https://doi.org/10.1007/978-981-13-6264-4_71

location of the network vulnerabilities is pointed. In [5], a neural network-based approach is presented for identifying the computer users to control the computer security, where three types of computer networks are taken into the consideration including a feedforward computer network, a sum of products network, and a corresponding hybrid network. Similarly, an online neural network based method is developed in [6], where the identification system consists of two phases with each corresponding to the investigation phase and the implementation phase. In [7], the security life of computer network is divided into two phases, where the first phase consists of security planning, hardware implementation, and protocol certification yet the second phase denotes the operational processing accidents. In [8], the security of the computer access to networks was first proposed to address in the year 1978, where a couple of the dedicated computers are proposed to store the keys and thus to solve the management of the keys. In [9], a novel approach was proposed in the year 1990 for the security of computer access, where the time between the two keystrokes entered by the potential users is proposed by using the pattern recognition method to secure the computer access security. All of these results are placed on the traditional computer security instead of involving wireless communications.

As to the computer security over the wireless communications, the authors in [10] introduce the wireless communications as the baseband interface over the security requirements. In [11], the security problem of the short-range wireless communications between two wireless devices is addressed since the wireless channel is probably to eavesdrop. Hence, a pairing scheme for initializing the secured wireless communications is presented. In [12], a self-learning mechanism is proposed to coordinate a couple of detection agents at each access point (AP) and thus to carry out the security detection in wireless LAN networks. In [13], the security service is address for the millimeter wave computer networks.

In this paper, we address the security of laptop access to the wireless networks. Two algorithms are developed to emphasize, respectively, the data throughput and the energy efficiency with one eavesdropper existence. All connections are wireless rather than wired. It is assumed that the existence of the eavesdropper is already known, yet the information related to this eavesdropper is not perfect. Both the eavesdrop capability and the desired capability of the legitimate laptop are taken into account.

2 System Model

As shown in Fig. 1, the studied system consists of $M \geq 2$ legitimate laptops which try to access the wireless LAN networks. Yet there is an eavesdropper who is active in this network. In this paper, it is assumed that each channel resource block such as slots, antennas, and frequency bands is only occupied by single legitimate laptop. The focus is placed on the one channel resource block to study the security of laptop access to wireless networks in this paper. All the noise are additive and independently distributed with each other. The goal of this paper is to design a method to select a proper laptop from all the legitimate

laptop candidates, where one eavesdropper exists on this channel resource block of wireless networks. All the laptop candidates are randomly distributed within the coverage of the access point (AP) in wireless LAN networks.

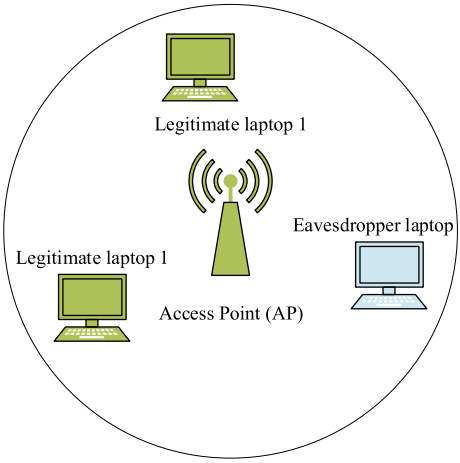


Fig. 1. Secured access to wireless LAN for one legitimate laptop with the existence of one eavesdropper laptop.

Considering the instantaneous capability of the eavesdropper together with the varying channel fading of the legitimate laptop candidate, the proper laptop shall be selected from all the legitimate laptop candidates to improve the data throughput or the energy efficiency as much as possible.

3 Protocol Design for Legitimate Laptop Selection with One Eavesdropper Existence

In this section, we propose two algorithms with each corresponding to the best legitimate laptop selection with the existence of one eavesdropper within the same wireless local area networks (LAN). The objective of each proposed algorithm is, respectively, to improve the secured throughput and the energy efficiency as much as possible. Considering the large scale channel fading, the path loss mainly depends on the distance. Besides, the small scale channel fading of both the eavesdropper and the legitimate laptop varies all the time. Hence, all these factors will be taken into the consideration for designing the legitimate laptop selection protocol.

Concerning the calculation of the involved SNR in SNR_m and SNR_e, the received signal at the mth legitimate laptop is given by

$$y_m = \sqrt{d_m^{-\alpha_m} p_t} h_m x + n_m, \quad (m = 1, 2, \cdots, M) \tag{2}$$

Algorithm 1. Throughput maximization based legitimate laptop selection with the existence of one eavesdropper in WLAN networks

Set the number of the legitimate laptop candidate $m = 1$, calculate the secured throughput of the first laptop $R_{s,0}$ according to the following expression,

$$R_{s,m} = \log_2(1 + \mathrm{SNR_m}) - \log_2(1 + \mathrm{SNR_e}), \tag{1}$$

where $R_{s,m}$ denotes the secured throughput corresponding to the mth legitimate laptop, $\mathrm{SNR_m}$ denotes the signal to noise ratio (SNR) of the mth legitimate laptop, $\mathrm{SNR_e}$ denotes the SNR of the eavesdropper.

repeat

set $m = m + 1$,$R_{s,max} = \max\{R_{s,max}, R_{s,m}\}$,$m_{max} = \ell\{R_{s,max}\}$, where $\ell\{\cdot\}$ denotes the subscript of $\{\cdot\}$ that corresponds to the $\ell\{\cdot\}$th legitimate laptop.

until m=M

where d_m is the distance for the wireless communications from the mth legitimate laptop candidate to the access point (AP), α_m is the path loss factor corresponding to the mth legitimate laptop which is typically in [3,5], h_m is the wireless channel that is usually distribute with Gaussian model $\mathbb{CN}(0, 1)$, x is the transmit signal from AP with the unit power, n_m is the noise received at the mth legitimate laptop which is usually treated as additive white Gaussian noise with the distribution $\mathbb{CN}(0, \sigma_m^2)$. Thus, SNR_m is described as $\mathrm{SNR}_m = \frac{p_t d_m^{-\alpha_m} |h_m|^2}{\sigma_m^2}$. Similarly, SNR_e is calculated according to $\mathrm{SNR}_e = \frac{p_t d_e^{-\alpha_e} |h_e|^2}{\sigma_e^2}$, where α_e is the path loss factor of the eavesdropper, σ_e^2 is the covariance of the noise at the eavesdropper, h_e denotes the wireless channel from AP to eavesdropper, d_e is the distance between AP and eavesdropper. The developed algorithm is presented in Algorithm 1 for selecting the best legitimate laptop for the secured throughput maximization.

Remark 1. The proposed Algorithm 1 addresses the instantaneous throughput of both the legitimate laptop and the eavesdropper, where the distance between the legitimate laptop and the eavesdropper involves in the calculation of SNR_m and SNR_e in terms of the path loss channel fading.

The proposed Algorithm 1 emphasizes the secured throughput in WLAN. Yet, the energy efficiency is not addressed which measures the throughput in the unit power consumption. Therefore, we design another algorithm to select the proper laptop with the existence of one eavesdropper to achieve the best energy efficiency for the secured wireless networks access.

Remark 2. The proposed Algorithm 2 addresses the energy efficiency of the secured wireless access over WLAN networks. In different regimes of transmit power p_t, the energy efficiency varies significantly even with the fixed legitimate laptop and the fixed eavesdropper in the same WLAN networks.

The proposed Algorithm 2 emphasizes the impact of the transmit power value over the instantaneous energy efficiency of the legitimate laptop for secured

Algorithm 2. Energy efficient selection of legitimate laptop with the existence of one eavesdropper in WLAN networks

Set the number of the legitimate laptop candidate $m = 1$, calculate the secured energy efficiency of the first laptop $\phi_{s,0}$ according to the following expression,

$$\phi_{s,m} = \frac{\log_2(1 + \mathrm{SNR_m})}{p_t} - \frac{\log_2(1 + \mathrm{SNR_e})}{p_t}, \tag{3}$$

where p_t is the transmit power from the access point (AP) as transmitter in WLAN networks.

repeat

set $m = m + 1$,$\phi_{s,max} = \max\{\phi_{s,max}, \phi_{s,m}\}$,$m_{max} = \ell\{\phi_{s,max}\}$.

until m=M

wireless access, which is the key parameter in the next generation of wireless LAN and 5G. This is different from the existing methods by minimizing the power under the constraint of predetermined quality of service since our objective in the proposed Algorithm 2 is to improve the number of bits in one second in one Hertz within one Joule. In fact, the energy efficiency varies substantially in different regimes of transmit power. The negative effect from the eavesdropper involves in designing the energy efficient selection of the proper legitimate laptop.

Remark 3. The proposed two algorithms for the proper legitimate laptop selection fully address the existence of the eavesdropper in the same WLAN networks from the wireless access point of view. However, the objectives in these two algorithms are different, where Algorithm 1 addresses the secured throughput yet Algorithm 2 emphasizes the secured energy efficiency for wireless access of the legitimate laptop to computer networks. These two algorithms can be applied to the access of WLAN, respectively, for throughput efficient and energy efficient networks.

4 Numerical Simulations with Discussions

In this section, the proposed two algorithms are numerically verified by Monte Carlo method, where the number of recycles in Monte Carlo is set to $5,000$. All the random variables are set to $\mathbb{CN}(0, 1)$. The path loss factor is set to $\alpha_m = \alpha_e = 2$. The radius of the WLAN networks is set to $r = 2$ km. The transmit power p_t varies from 0 dB to 30 dB. The number of the legitimate laptop candidates is set to $M = 10, 30, 50$. Every legitimate laptop is randomly distributed in the circle of WLAN. The distance between AP and the eavesdropper is randomly determined every time of channel realization.

Figure 2 plots the secured throughput realized by our proposed Algorithm 1. In various regimes of the transmit power, the proposed laptop access protocol has a substantial gain compared to the existing method, where the random laptop selection is chosen as the benchmark method. Moreover, with the increase of the number of the legitimate laptop candidates, the secured throughput grows for any given value of the transmit power. This is because the diversity of the laptop

selection gain is utilized. Similarly, Fig. 3 presents the energy efficiency achieved by the proposed Algorithm 2, which is the best from the energy efficiency point of view since the objective in Algorithm 2 is to maximize the secured energy efficiency.

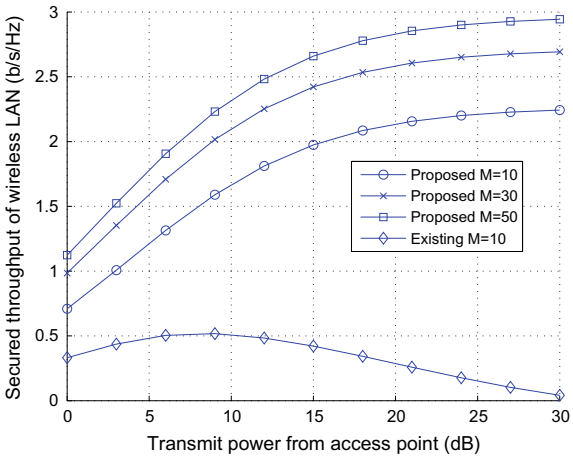


Fig. 2. Secured throughput achieved by proposed Algorithm 1

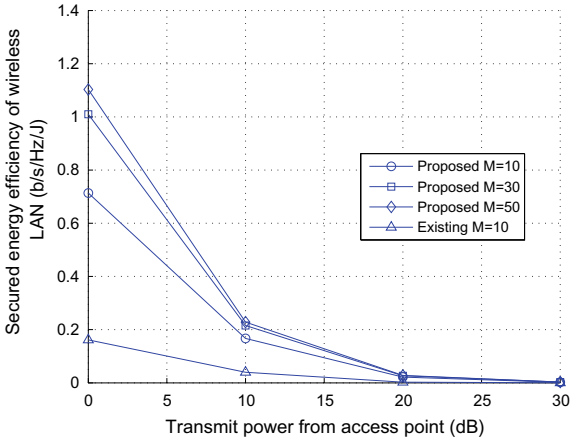


Fig. 3. Secured energy efficiency achieved by proposed Algorithm 2

5 Conclusion

Two efficient laptop scheduling algorithms are proposed for the selection of a couple of legitimate laptops in wireless LAN networks with one eavesdropper laptop. These two algorithms address the information leakage from the eavesdropper yet emphasizes, respectively, the throughput and the energy efficiency of the secured wireless access in wireless LAN networks. The distance of the wireless access, respectively, for both the legitimate laptop and the eavesdropper laptop is taken into consideration. By using the proposed algorithms, the throughput of the legitimate access to wireless LAN is the best achieved by Algorithm 1. Similarly, the proposed Algorithm 2 achieves the best energy efficiency for the secured access to this network. The method of designing the efficient secured access to networks can be applied to other scenarios such as cellular systems.

References

1. Landwehr, C.E.: The best available technologies for computer security. Computer **16**, 86–100 (1983)
2. Summers, R.C.: An overview of computer security. IBM Syst. J. **23**, 309–325 (1984)
3. Abrams, M.D., Jeng, A.B.: Network security: protocol reference model and the trusted computer system evaluation criteria. IEEE Netw. **1**, 24–33 (1987)
4. Andrews, M., Whittaker, J.A.: Computer security. IEEE Secur. Priv. **2**, 68–71 (2004)
5. Obaidat, M.S., Macchairolo, D.T.: A multilayer neural network system for computer access security. IEEE Trans. Syst. Man Cybern. **24**, 806–813 (1994)
6. Obaidat, M.S., Macchiarolo, D.T.: An online neural network system for computer access security. IEEE Trans. Ind. Electron. **40**, 235–242 (1993)
7. Pierson, L.G., Witzke, E.L.: A security methodology for computer networks. AT&T Tech. J. **67**(3), 28–36 (1988)
8. Branstad, D.: Security of computer communication. IEEE Commun. Soc. Mag. **16**, 33–40 (1978)
9. Bleha, S., Slivinsky, C., Hussien, B.: Computer-access security systems using keystroke dynamics. IEEE Trans. Pattern Anal. Mach. Intell. **12**, 1217–1222 (1990)
10. Mueck, M., Ivanov, V., Choi, S., Kim, J., Ahn, C., Yang, Y., Baldini, G., Piipponen, A.: Future of wireless communication: RadioApps and related security and radio computer framework. IEEE Wirel. Commun. **19**, 9–16 (2012)
11. Gallego, A., Saxena, N., Voris, J.: Playful security: a computer game for secure wireless device pairing. In: 2011 16th International Conference on Computer Games (CGAMES), Louisville, KY, pp. 177–184 (2011)
12. Chang, Z., He, Q., Zhao, P.-Y., Liu, J.-S.: A wireless-LAN security model based on computer immunology. In: 2009 International Conference on Machine Learning and Cybernetics, Baoding, pp. 2054–2057 (2009)
13. Minin, I.V., Minin, O.V.: Information security services in the millimeter-wave computer network with wireless access. In: 2005 5th International Conference on Microwave Electronics: Measurement, Identification, Applications, pp. 154–157 (2005)

An Interruption-Resilient Service Priority-Oriented TDMA Scheme for NAANET

Cong Yang and Zhihua Yang(✉)

Shenzhen Graduate School, Communications Engineering Research Center, Harbin Institute of Technology, Shenzhen, China
yangzhihua@hit.edu.cn, 2018yangcong@gmail.com

Abstract. This paper proposes a novel dynamic time-division multiple access (TDMA) protocol, which can tolerate stochastic channel interruptions for UAV communication in low-altitude environments. In particular, the proposed protocol dynamically adjusts the allocation of slots by predicting the probability of channel interruption. As a result, this method could improve the utilization of time slots through first guaranteeing the transmission of high-priority services. Simulation results show that compared with the existing dynamic TDMA protocol, the novel slot allocation algorithm based on the channel outage probability reduces the packet loss rate of the service and shortens the average delay of system.

Keywords: Channel interruption · TDMA · Priority · Outage probability

1 Introduction

A UAV Ad Hoc Network (UAANET) is a dynamic self-organizing network system composed of UAV network nodes with autonomously time-varying network topology [1]. In a low-altitude and high-dynamic environment, terrain conditions will have certain impacts on the channel access of UAANET. In particular, obstacles such as trees and buildings will hinder the connection of UAV communication links.

At present, typical UAANET multi-access channel (MAC) protocol mainly includes contention-based MAC protocol and scheduling-based TDMA protocol. The competition-based MAC protocol mainly includes random competition and reservation competition. Due to the high transmission collision probability of the random contention-based MAC protocol [2], the transmission success ratio and the channel utilization performance are severely degraded with the growth of the traffic volume, which is not suitable for a low-altitude and high-dynamic environment. Based on the IEEE 802.11 DCF subscription-based MAC protocol [3], the core mechanism is CSMA/CA. It is required to monitor whether the channel is idle. If the channel is randomly interrupted, data transmission cannot be guaranteed, which has a great impact on the protocol performance. While the fixed TDMA protocol [4] based on the scheduling class is added with the service priority [5, 6] and the competition mechanism, the performance of the protocol has been improved, but when the channel is randomly interrupted, the network has lower quality of service (QoS).

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 605–613, 2019.
https://doi.org/10.1007/978-981-13-6264-4_72

Based on the dynamic and prioritized TDMA protocol, this paper proposes a slot allocation algorithm with random interrupt tolerance, which improves the shortcomings of traditional protocols for handling channel random interruption. The algorithm predicts the probability of the next frame channel interruption based on the communication quality of the current environment channel, and adjusts the time slot assignment of the next frame according to the prediction value. Simulation results show that this novel algorithm improves channel utilization and reduces the packet loss rate.

The structure of this paper is as follows: Sect. 2 describes the design process of the protocol algorithm. Section 3 describes the random interrupt feedback algorithm. Section 4 established a network model and derived a performance index formula. Finally, introduce the experimental results and conclusions.

2 Protocol Algorithm Design

2.1 Queue Scheduling Model

This article according to the specific TDMA protocol slot scheduling scheme designed by the corresponding queue scheduling algorithm, ignore the complicated MAC basic protocol control flow and time frame structure. The queue scheduling is mainly divided into two parts: packets are enqueued for transmission classification management and dequeued by the schedule algorithm. Among them, the classification management of enqueued packets is mainly to prioritize the data and store them in the corresponding priority queues. The scheduling algorithm mainly embodies the idea of time-sharing scheduling, and performs priority scheduling in fixed-assigned time slots, dynamically adjusts the transmission of service during dynamical slots. Assume UAV has a transmitter and multiple receivers, and both can switch and work quickly on multiple independent sub-frequencies (channels). The specific queue scheduling algorithm is shown in Fig. 1.

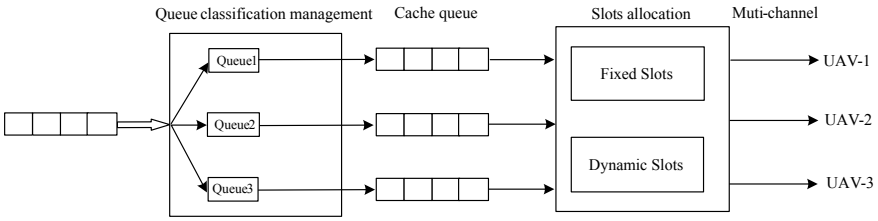


Fig. 1. Priority queue scheduling model

2.2 Priority Queue Classification Management

The application layer request of each node is sent to the MAC layer and mapped to a specific priority packet and stored in the MAC layer request buffer queue. The buffer queue is actually a priority queue group. Corresponding to each priority packet, there is a buffer queue. For example, if there are 1–3 levels with a total of three priorities, then

there are at least three priority queue groups as buffer queues. Data of priority level i $(1 \leq i \leq 3)$ package will be sent to queue i for waiting.

The queue buffer area is divided into buffer A and buffer B, the buffer B is a reserved area, in which the capacity of buffer A is greater than buffer B. The priority queue packet prior to enter buffer A, and the queue is assigned to the fixed time slot according to the priority level, then the remains queue is assigned to the dynamic time slot according to the priority level. If the transmission is interrupted, the priority queue cannot be sent in this frame, and the queue is passed to buffer B. When the priority queue in buffer B is sent depends on the channel connection. If the channel is not restored for a long time, remove the priority queue from the buffer B. If the buffer B queue is full, the lower priority queue is preferentially removed (Fig. 2).

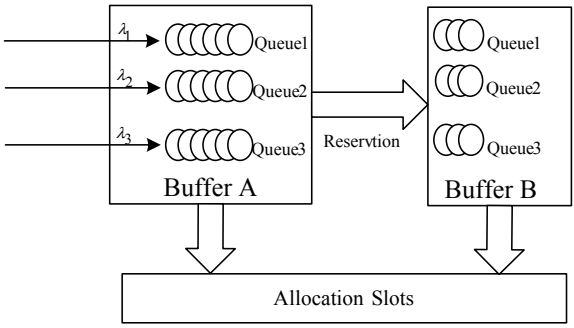


Fig. 2. Priority queue buffer allocation

2.3 Dynamic Slot Allocation

Assuming that one of the UAVs serves as the master node to communicate with the other three UAVs, three channels need to be established between them. Assume that the three types of priority services are high-priority service $A_1B_1C_1$, medium priority service $A_2B_2C_2$, and low-priority service $A_3B_3C_3$. Ignoring the structure of the preamble slot in the frame, the length of each type of priority data packet is fixed, and M is the preset value of dynamic time slots allocation, the master node is needed to schedule dynamic reserved time slots, and the statistics node at the end of the frame is needed to statistics the time slots of success data transmission (Fig. 3).

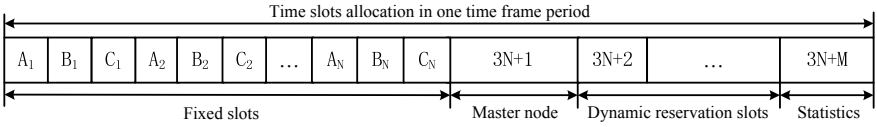


Fig. 3. The frame structure of multichannel TDMA protocol

The dissertation algorithm assigns resources to each priority in the order of serial number. First, set the number of fixed timeslots allocation according to the priority requests of different nodes. Then, according to the remaining number of requests, a

dynamic time slot is allocated for each node. The size of the preset value M can be adjusted according to the service conditions. The total number of dynamic timeslots allocation decided by the dynamic requests but does not exceed twice the value of M in principle. The flow of the dynamic time slot assignment algorithm is as follows:

Algorithm 1: Dynamic Slots Allocation

```
1: //Initialization
2: //Count the total number of all requests.
3: if (total number of requests > total number of scheduled data slots )
4:   if (total number of high priority requests > total number of scheduled data slots)
5:     // Expand the number of scheduled data slots and allocate high priority requests.
6:     // Adjust other priority allocation requests.
7:   else
8:     // Assign high priority requests first. The remaining data slots are assigned to others.
9:     // The number of scheduled data slots remains unchanged;
10: else
11:   // Calculate the remaining amount slots.
12:   //All requests are allocated for data slots, and the remaining data slots are deleted;
13:  end
```

3 Interruption Probability Feedback Algorithms

This paper mainly considers UAANET in low-altitude environment. Terrain conditions will have a certain impact on the coverage and propagation of UAV signals. The communication link between UAVs may encounter channel fading due to the blockage of the building or occurence channel random interruptions as the influence of mountains. Assuming there are N nodes in a UAANET, $i, j \in N$ represent nodes of the network. When the SNR between nodes is greater than a certain threshold γ, we think of node i and node j can communicate directly.

$$SNR(i,j) = \frac{P_i}{L_b(i,j)N_e} \geq \gamma \tag{1}$$

where P_i is the transmit power of node i, $L_b(i, j)$ is the path loss between node i and node j, and N_e is the noise influence parameter caused by the current environmental factors.

Dividing a frame time T_{frame} into N segments, the corresponding time interval is $[t_a, t_b]$, and T_i represents time interval $[t_{ai}, t_{bi}]$, where $i = 0, 1, \ldots, N-1$, $T_i = T/N$. Let $A_i = 1$ denote the event of SNR is greater than or equal to threshold γ at time t_{ai}, then $A_i = 0$ denote the event of SNR is less than threshold γ at time t_{ai}. Assume that the probability of an interruption occurring in the transmission of data per frame between node i and node j is P. We can deduce the probability value P in (2):

$$P = \frac{\sum_{i=0}^{N-1} A_i}{N} \tag{2}$$

During data transmission, nodes may experience two states. One is that the channel is interrupted during transmission, and the other is that the channel remains connected. Note that, event B is the channel interruption of the UAV node in a certain frame, and the probability of channel interruption in the transmission of the Xth frame is $p(B_X)$, and the distribution of the outage probability as follows Table 1:

Table 1. Outage probability distribution

Frame	1	2	3	…	X
P	$p(B_1)$	$p(B_2)$	$p(B_3)$	…	$p(B_X)$

The slots allocation of the node is adjusted by predicting the value of the outage probability. If the outage of the channel is a small probability event, the slots are still allocated for the node; if the outage probability of the node is large, then pause the slots allocation of the next frame, after the node's outage probability is reduced a certain value, resumes the allocation; if the node's outage of channel is medium probability, then combined with outage probability of the previous frames to make judgment. Through probability, feedback strategy reduces the probability of failure in data transmission and improves the utilization of the channel. If a certain UAV node is not detected for a long time, the node is considered to have been destroyed; if the node information is received, then the node accesses the channel in the next frame. The channel outage probability algorithm is as follows:

Algorithm 2: Outage Probability Feedback Algorithm

1: //Initialization
2: **if** ($p(B) < a$)
3: **if** (The number of interruptions in a node in consecutive N frames <=1)
4: // Timeslot allocation strategy unchanged
5: **else if** (The number of interruptions in a node in consecutive N frames >1)
6: //Adjust the slot allocation strategy
7: **if** ($a<p(B) < b$)
8: //Count the number of interrupts occurring in consecutive N frames
9: **if** (The number of interruptions in a node in consecutive N frames <=1)
10: // Timeslot allocation strategy unchanged
11: **else if** ($\sum B_{i-1} B_i > 0$ && $\sum B < N/2$)
12: **else**
13: //Regard the node as corrupt and clear the data in the cache
14: **if** ($p(B) > b$)
15: // Suspend slot allocation, send data when the channel outage probability is detected to be less than b
16: **end**

4 Performance Analysis

Usually, the rate of a UAV node generates data packets obeys a Poisson distribution with parameter λ, and the MAC layer processes these data obeys a uniform distribution with parameter μ. This behavior is similar to the model of queuing network [7] with the external arrival rate λ packets/s and service rate μ packets/s. Hence, the queuing model is suitable for the UAVs to examine the performance parameters.

In order to facilitate the analysis of queuing network model for UAANET, some popular symbols are defined in Table 2.

Table 2. Notations

Symbols	Parameter specification
λ_i	The arrival rate of packets with priority i
μ_i	The service rate of packets with priority i
t_i.	The channel interruption rate
W_i	Waiting queue time
W_d	Average delay time
W_s	Average service time
N	Number of nodes

Suppose that packets with different priorities are transmitted to different queues. The number of nodes is N, the data packets obey a Poisson distribution with rates λ_i, and i denotes its priority level. The process of a data packet occupying a time slot and sending it in a specified time slot can be seen as the customer queuing to receive the service process. Assume that the processing rate of each priority is μ_i. From Little Theorem, the average service time is

$$\overline{S_i} = E(S_i) = \frac{1}{\mu_i} \tag{3}$$

Denotes $\rho_i = \lambda_i/\mu_i$ ($1 \leq i \leq m$). Since the first to Nth data packets are Poisson streams that are independent of each other, the packets arriving belong to the ith priority is λ_i/λ. Thus, the average service time for the system is

$$\overline{S} = E(S) = \sum_{i=1}^{m} \frac{\lambda_i}{\lambda} E(S_i) = \frac{1}{\lambda} \sum_{i=1}^{m} \rho_i \tag{4}$$

The delay of the service with priority c_i ($1 \leq i \leq m$) includes three parts: the remaining processing time required for the high-priority service, the waiting time for this service being processed and the interruption time of this service.

If the service process is interrupted, assuming that the interruption interval obeys the Poisson distribution f_i, the duration obeys the Poisson distribution g_i, then the average interruption time is

$$\overline{S_{\text{off}}} = E(S_{off}) = \sum_{i=1}^{m} \frac{\lambda_i}{\lambda} E(S_{i\_off}) = \sum_{i=1}^{m} \frac{\lambda_i \mu_i}{\lambda f_i g_i} \tag{5}$$

The number of all packets with priority i is M, and the processing time of each packet is S_i, ignore the other time consumption. Average queuing time of nodes with priority l is as in (6):

$$W_{ql} = \frac{1}{\left(1 - \sum_{i=1}^{l} \rho_i\right)\left(1 - \sum_{i=1}^{l-1} \rho_i\right)} \cdot \frac{\lambda E(S^2)}{2} \tag{6}$$

From the formulas above, we can deduce the delay for a packet with priority i is the sum of queuing delay and processing delay in (7):

$$D_i = S_{i\_off} + W_{qi} \tag{7}$$

Assume that the length of send packet with priority i is L_{si}, the length of received packet with the priority i is L_{ri}, then the loss rate with priority i is

$$L_{i\_loss} = \frac{L_{si} - L_{ri}}{L_{si}} \tag{8}$$

Assume that the simulation time is T, then the throughput with priority i is

$$T_{i\_throughput} = \frac{L_{si} - L_{ri}}{\mathrm{T}} \tag{9}$$

5 Simulation Results

In this section, we will make experimental simulations for evaluations on the proposed algorithm. For the simplicity of analysis, without loss of generality, we proposed a priority-based dynamic packet reservation for TDMA wireless networks, which consider the outage probability feedback algorithm. The parameters of the simulation are defined as follows (Fig. 4):

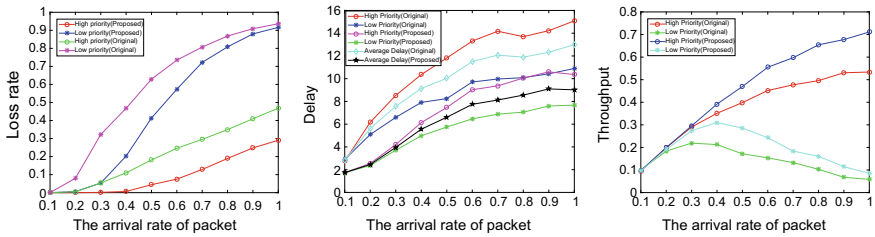


Fig. 4. Data packet loss ratio for each priority and the delay of each priority and throughput for each priority

Table 3. Parameters of simulation model

Parameter	Parameter specification	Value
T	Simulation time	100000 s
μ	Service rate	1
λ_1	HP rate	0.1–1
λ_2	LP rate	0.1–1
p	Frame outage probability	*Poisson* (0.1)

Compared with low-priority services, the performance of high-priority services has clearly improved. As can be seen from the Fig. 4, the packet loss rate of high-priority services is significantly better than low-priority under heavy-load and medium-load conditions. Under the outage probability feedback algorithm, the packet loss rate of high-priority services is only half of that of the original method at heavy load and medium load, while the packet loss rate of low-priority services increases slightly with the packet's arrival rate increases. The average delay of the system is clearly shortened. The throughput of high-priority and low-priority are both improved by the new method Table 3.

When the buffer capacity is fixed, the change of the threshold value has a significant impact on the QoS of the system. Assume that the high-priority threshold $\gamma_1 = 10$, the low-priority threshold γ_2 increases from 1 to 10, and the other parameters are set as $\lambda_1 = \lambda_2 = 1$, $\mu = 2$.

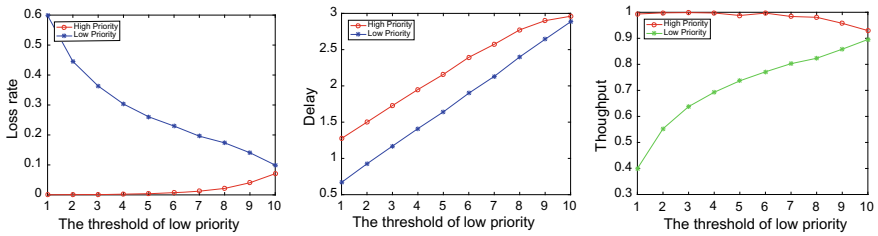


Fig. 5. Impact of threshold changes on system service performance

As can be seen from Fig. 5, as the threshold increases, the loss rate of low-priority queues decreases, while the loss rate of high-priority queues increases. At the same time, the average delay of both increases. The increase of the threshold has less impact on the high-priority throughput, while the low-priority throughput has an increasing trend. When the threshold is 10 (both priorities share the buffer completely), the loss rate, average delay, and throughput are very close. The average delay of high-priority is greater than the average delay of low-priority, because the high-priority queues occupy larger buffers to receive more queues. Threshold values have a great influence on QoS. According to different traffic characteristics, an appropriate threshold value can be selected to improve the utilization of system resources while protecting high-priority queues as much as possible.

6 Conclusion

In this paper, we focused on the study the impact of burst channel interrupts to the transmission services. Based on the priority-based TDMA protocol, the new slot allocation is improved to meet the retransmissions of the interrupt node. For the interrupt node service, an outage probability feedback algorithm is proposed. The simulation results show that the packet loss rate based on the outage probability feedback algorithm is smaller than the original priority dynamic TDMA protocol, and the average delay is also shortened, and the throughput of system is clearly improved.

Acknowledgments. This work was supported by the Basic Research Program of Shenzhen city under Grants (JCYJ20170413110004682, JCYJ20150403161923521, and JCYJ20150625142 543458).

References

1. Song, L., Huang, T. l.: A summary of key technologies of ad hoc networks with UAV node. In: 2010 International Conference on Intelligent Computing and Integrated Systems, Guilin, pp. 944–949 (2010)
2. Wang, W., Dong, C., Zhu, S., Wang, H.: DFRA: demodulation-free random access for UAV ad hoc networks. In: 2017 IEEE International Conference on Communications (ICC), Paris, pp. 1–6 (2017)
3. Xu, C., He, J.: Performance analysis of IEEE 802.11 cognitive radio ad hoc networks. In: 2015 IEEE/CIC International Conference on Communications in China (ICCC), Shenzhen, pp. 1–6 (2015)
4. Gronkvist, J., Karlsson, J., Sterner, U., et al.: Adaptation delay and its impact on application performance for TDMA ad hoc networks. In: The 11th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net), pp. 55–60. IEEE Conference Publications (2012)
5. Ke, P., Li, W., Wu, Y.: Research on multiple services scheduling based on priority-queuing model. In: 2014 IEEE 12th International Conference on Dependable, Autonomic and Secure Computing, Dalian, pp. 372–376 (2014)
6. Das, D., Abouzeid, A.A.: Delay analysis of multihop cognitive radio networks using network of virtual priority queues. In: 2014 IEEE Wireless Communications and Networking Conference (WCNC), Istanbul, pp. 1514–1519 (2014)
7. Van Mao, N., Son, V.Q.: Applying queuing theory to evaluate performance of cluster wireless sensor networks. In: 2015 International Conference on Advanced Technologies for Communications (ATC), Ho Chi Minh City, pp. 501–506 (2015)

Dimensioning and Layout Planning of Integrated VANET-WSN for Intelligent Transportation

Wei Zhang, Bin Lin(✉), Cunqian Yu, Ying Wang, and Changxuan Gao

College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China
binlin@dlmu.edu.cn

Abstract. The future development of Intelligent Transportation relies heavily on a "smart" road as the fundamental infrastructure. The Integrated Vehicular Ad Hoc Network (VANET)–Wireless Sensor Network (WSN) has provided a promising network architecture solution for building smart roads. Besides, WSN can enhance the road safety and provide other roadside services, monitor road conditions, and report to drivers by using sensors, which are also deployed at roadside together with the Road Side Units (RSUs) in VANETs. This paper investigates the optimal roadside infrastructure Dimensioning and Layout Planning (DLP) problem under the Integrated VANET-WSN architecture to achieve total coverage along the road by considering communication constraints, coverage requirement, connectivity, and network characteristic between the RSUs and the sensors. We formulate the problem as an Integer Linear Programming (ILP) and solve it by Gurobi. Simulation results provide an optimal layout solution, and the results show that the proposed architecture is cost-effective and scalable in practice.

Keywords: Integrated VANET-WSN network · DLP · Cost-effective · ILP

1 Introduction

Smart roads, as one of the most significant ITS infrastructure, make the road and highway transportation more efficient and safer [1], enable innovative control of traffic, alleviate traffic congestion, and better navigation and driving conditions for vehicles with guidance in the transport networks [2, 3]. It is crucial for the transportation-related messages to be delivered correctly and effectively to the drivers. Based on wireless communication technology, Vehicular Ad Hoc Networks (VANETs) can provide the information sharing platform [4]. There are many sensor nodes in WSN that can effectively detect environmental events, transmit data to each other through multi-hop

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 614–622, 2019.
https://doi.org/10.1007/978-981-13-6264-4_73

communication and finally report it to the Base Station (BS). In this paper, we focus on the sensitive areas that are dangerous for driving, especially the icy roads in winter and the low-lying roads with water. In the VANET-WSN Network, the roadside sensor nodes can collect these events and pass directly to the passing vehicles.

The concept of Integrated VANET-WSN network has drawn wide attention recently. The nodes in the WSN are generally highly miniaturized, mostly static, with limited resources and energies, have good sensing capabilities [5]. In contrast, VANET has a dynamic topology. By integrating VANET and WSN, the two types of networks can complement each other, enhance road intelligence [6]. We investigate the deployment problem as a Dimensioning and Layout Planning (DLP) problem, select BSs, RSUs, and sensors from candidate points, join link loss into consideration and advance the concept of "flow" between the RSU and Test Points (TPs) to deal with coverage and connectivity. As shown in Fig. 1, vehicles passed by and other vehicles in a wider area via VANET will receive traffic information from the network and reduces traffic accidents.

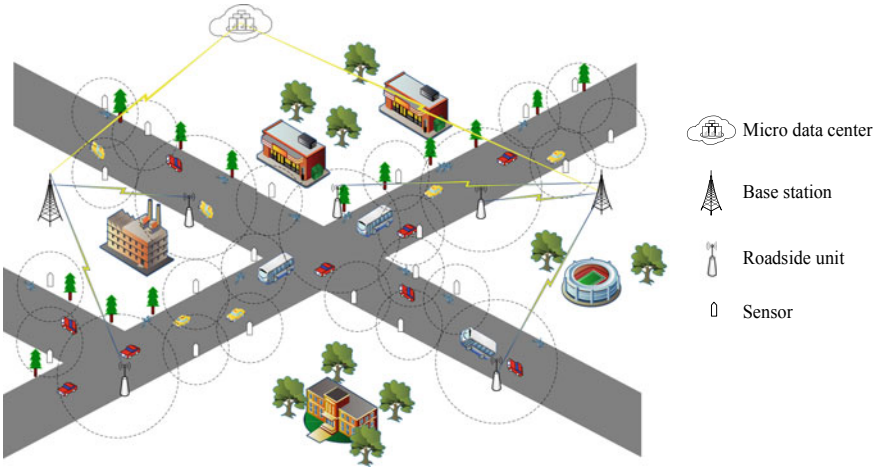


Fig. 1. The architecture of integrated VANET-WSN

2 Network Architecture and Problem Formulation

2.1 Network Architecture

In this section, we discuss the architecture of the Integrated VANET-WSN network. The tree-based architecture is described in Fig. 2. The VANET-WSN architecture consists of TPs, sensors, RSUs, BSs, and Micro data center. As show in Fig. 2, the solid line mean RSU use wired mode to communicate with BS, and the dotted line represents wireless communication. The tree-based architecture is described in Fig. 2.

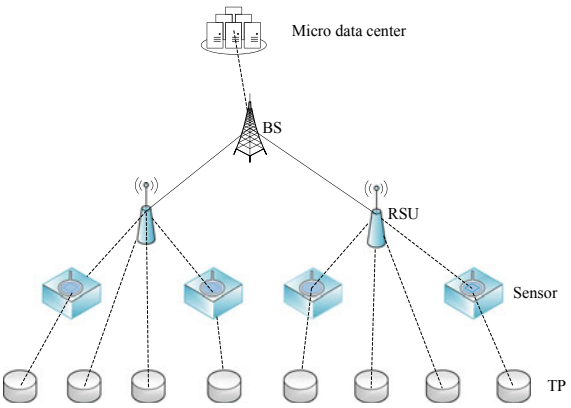


Fig. 2. The tree-based architecture

- By representing the road as evenly distributed points, these points are called Test Points (TPs). Covering the entire road by covering all the TPs, so that low-lying sections with frequent water-logging and icing-slippery road conditions are fully monitored to improve transportation safety on the city roads.
- Sensors are deployed along the roadsides and persevered in detecting road conditions and reporting them. They directly deliver what they obtain in the sensitive areas to the vehicles passing by and spread to other vehicles in a wider area via VANET.
- Road Side Units (RSUs) receive the messages from the sensors within the communicated range, and then RSUs could discover and communicate with vehicles. RSUs work as access points, so it is necessary to place RSUs in the appropriate locations.
- Base Stations (BSs) are used to collect traffic information from the RSUs and provide communications with the Internet.
- Micro Data Center (MicroDC) is the root of the tree topology within the architecture. It is used to store and extensively analysis all the traffic information and road environment information collected by the BS.

2.2 Problem Formulation

We consider the network as a directed graph $G(\Omega, E)$, where E is the set of directed edge and the node set Ω includes BSs set Ω_{BS}, RSUs set Ω_{RSU}, sensors set Ω_{sensor}, and TPs set Ω_{TP}, set $\Omega = \Omega_{BS} \cup \Omega_{RSU} \cup \Omega_{sensor} \cup \Omega_{TP}$. And the cost of a BS C^B, the cost of a RSU C^R and the cost of a sensor C^S.

Given the locations of the MicroDC and TPs, the candidate points for the BSs, RSUs, and sensors, then select the optimal entities from the various candidate points. The constraints are included coverage, connection, topology, capacity, transmission distance, RSUs, and sensors candidate points are colocated and the total path loss traversing every "link flow" is less than the minimum threshold.

2.3 Variables

Four sets of decision variables are used to define the DLP problem. Specifically, we define location matrix, path selection matrix, coverage matrix, and flow matrix.

(1) BS, RSU, and sensor's location matrix is set to $A = (a_i)_{1\times i}$, $B = (b_j)_{1\times j}$, $C = (c_t)_{1\times t}$ and place different network element devices at different candidate nodes (PS_s).
If place the RSU node at the candidate point PS_i, $a_i = 1$; otherwise, $a_i = 0$.
If place the sensor node at the candidate point PS_j, $b_j = 1$; otherwise, $b_j = 0$.
If place the BS at the candidate point PS_t, $c_t = 1$; otherwise, $c_t = 1$.

(2) $E = (e_{ij})_{i\times j}$, $i \in \Omega_{sensor} \cup \Omega_{RSU} \cup \Omega_{BS} \cup \Omega_{DATA}, j \in \Omega_{TP} \cup \Omega_{sensor} \cup \Omega_{RSU}$ is path selection matrix, such that
If the path from node i to node j is selected, $e_{ij} = 1$; otherwise, $e_{ij} = 0$.

(3) Coverage of TPs in the perceptual area matrix, $D = (q_t)_{1\times t}$
If of the TPs can be covered in the sensing area, $q_t = 1$; otherwise, $q_t = 0$.

(4) The flow matrix F from RSUs to TPs $(f_{ij}^{mk})_{i\times j}$, $\forall i \in \Omega_{RSU\cup sensor}$, $\forall j \in \Omega_{sensor\cup TP}, \forall m \in \Omega_{RSU}, \forall j \in \Omega_{TP}$.
If e_{ij} is on the path from RSUs to TPs, $f_{ij}^{mk} = 1$; otherwise, $f_{ij}^{mk} = 0$.

Table 1 lists the parameters used for problem formulation.

Table 1. The parameters used for problem formulation

Parameters	Definition
{P}	The position of MicroDC
BS_COM	The communication distance of BSs
SN_SEN	The perceived distance of sensors
SN_COM	The communication distance of sensors
RSU_R	The coverage ratio of RSUs
D	The distance matrix. D = (d_{ij}), where d_{ij} is the Manhattan Distance between node i and node j
C^f	The price of per unit length fiber
H	The maximum number of sensors that the RSU can accommodate
k	The weight number that TPs are covered by the APs
$\partial_3, \partial_2, \partial_1, \partial$	Path loss between different types of nodes

2.3.1 Formulations

We formulate this DLP problem as an integer linear programming (ILP). The formulations are as follows.

Objective: minimize

$$C = C^R \sum\nolimits_{i\in\Omega_{RSU}} a_i + C^S \sum\nolimits_{j\in\Omega_{sensor}} b_j + C^B \sum\nolimits_{k\in\Omega_{BS}} c_k \tag{1}$$

The objective function (1) represents that the optimization goal is to get the minimum cost of the Integrated VANET-WSN network.

Subject to

$$\sum\nolimits_{j\in\{P\}} h_j = 1 \tag{2}$$

$$\sum\nolimits_{i\in\Omega_{BS}} e_{ij} \geq h_j, \forall j \in \{P\} \tag{3}$$

$$\sum\nolimits_{i\in\{P\}} e_{ij} = c_j, \forall j \in \Omega_{BS} \tag{4}$$

To guarantee that all applications, constraint (2) limits the number of MicroDC. In constraint (3), if a MicroDC is selected, there is at least one output from the MicroDC to the BS. Constraint (4) ensures that every BS is associated with the MicroDC.

$$\sum\nolimits_{j\in\Omega_{BS}} e_{ij} = a_i, \forall i \in \Omega_{RSU} \tag{5}$$

$$\sum\nolimits_{j\in\Omega_{RSU}} e_{ij} \geq c_i, \forall i \in \Omega_{BS} \tag{6}$$

In constraints (5) and (6), every BS can connect more than one child node, while the RSU must have one and only one parent node, respectively.

$$\sum\nolimits_{j\in\Omega_{sensor}\cup\Omega_{TP}} e_{ij} \geq a_{i,} \forall i \in \Omega_{RSU} \tag{7}$$

$$\sum\nolimits_{j\in\Omega_{sensor}} e_{ij} \geq H, \forall i \in \Omega_{RSU} \tag{8}$$

$$\sum\nolimits_{i\in\Omega_{RSU}\cup\Omega_{sensor}} e_{ij} = b_j, \forall j \in \Omega_{sensor}, i \neq j \tag{9}$$

$$\sum\nolimits_{j\in\Omega_{RSU}\cup\Omega_{TP}} e_{ij} \geq b_i, \forall i \in \Omega_{sensor}, i \neq j \tag{10}$$

$$\sum\nolimits_{i\in\Omega_{RSU}} a_i \geq 1 \tag{11}$$

$$a_i + b_i \leq 1, \forall i \in \Omega_{RSU}, j \in \Omega_{sensor}, i = j \tag{12}$$

In constraint (7), each RSU connects to multiple sensors and TPs. Constraint (8) ensures that the number of sensors which RSU contained should not exceed its capacity. Constraints (9) and (10) elaborate if a sensor node exists, there is exactly one input edge from RSU or other sensor to the sensor itself, and at least one output edge from the sensor itself to the other sensor or TP. Constraint (11) means at least one RSU is selected. In constraint (12), place one network element equipment for each candidate point. TPs should be covered by sensors and RSUs, and the coverage weight is 1.

$$e_{mt} \leq q_t, \forall m \in \Omega_{RSU} \cup \Omega_{sensor}, \forall t \in \Omega_{TP} \tag{13}$$

$$\sum\nolimits_{m \in \Omega_{RSU} \cup \Omega_{sensor}} z_{mn} \geq k, \forall n \in \Omega_{TP} \tag{14}$$

$$\frac{1}{N} \sum\nolimits_{n \in \Omega_{TP}} q_n \geq \rho_0 \times 100\% \tag{15}$$

In order to establish links between the nodes to make sure network connectivity, we have some constraints below. Transmission distance constraints in integrated VANET-WSN network are shown in constraints (19)–(23).

$$e_{ij \leq} c_i, \forall i \in \Omega_{BS}, j \in \Omega_{RSU} \tag{16}$$

$$e_{ij} \leq a_i, \forall i \in \Omega_{RSU} j \in \Omega_{sensor} \tag{17}$$

$$e_{ij} \leq b_i, \forall i \in \Omega_{sensor}, j \in \Omega_{sensor}, i \neq j \tag{18}$$

$$e_{ij} d_{ij} \leq SN\_SEN, \forall i \in \Omega_{sensor}, j \in \Omega_{TP} \tag{19}$$

$$e_{ij} d_{ij} \leq SN\_COM, \forall i, j \in \Omega_{sensor}, i \neq j \tag{20}$$

$$e_{ij} d_{ij} \leq RSU\_R, \forall i \in \Omega_{RSU} j \in \Omega_{sensor} \tag{21}$$

$$e_{ij} d_{ij} \leq RSU\_R, \forall i \in \Omega_{RSU}, j \in \Omega_{TP} \tag{22}$$

$$e_{ij} d_{ij} \leq BS\_COM, \forall i \in \Omega_{BS} j \in \Omega_{RSU} \tag{23}$$

We advance the concept of "flow" between the RSU and TPs. Considering the overall meaning of the Integrated VANET-WSN link, the link between nodes is subject to "link flow". Constraint (24) limits the maximal number of hops from RSU to TPs. Constraint (25) illustrates that the flow from node m to node k via e_{ij}. Constraint (26) donates each TP node should be included in a "link flow". Constraint (27) implies when an RSU node exists, it must have at least one outflow. Constraints (28) and (29) elaborate that the "link flow" between RSU and TPs is established. Constraint (30) indicates that sensor, as an intermediate point, establishes the relationship between the input stream and output stream. Constraint (31) is listed to guarantee that the total path loss traversing the "link flow" is below the minimum threshold.

$$\sum\nolimits_{\forall i \in \Omega_{RSU} \cup sensor, \forall j \in \Omega_{sensor} \cup TP} f_{ij}^{mk} \leq Hop_{\max}, \forall m \in \Omega_{RSU}, k \in \Omega_{TP} \tag{24}$$

$$f_{ij}^{mk} \leq e_{ij}, \forall i \in \Omega_{RSU} \cup sensor, \forall j \in \Omega_{sensor \cup TP}, \forall m \in \Omega_{RSU}, k \in \Omega_{TP} \tag{25}$$

$$\sum\nolimits_{m \in \Omega_{RSU}} \sum\nolimits_{i \in \Omega_{RSU} \cup \Omega_{sensor}} f_{ij}^{mk} = 1, \forall k \in \Omega_{TP}, j \in \Omega_{TP} \tag{26}$$

$$\sum_{m\in\Omega_{RSU,}}\sum_{j\in\Omega_{sensor}\cup\Omega_{TP}} f_{mj}^{mk} \geq a_m, \forall k \in \Omega_{TP} \tag{27}$$

$$f_{ij}^{mk} \leq a_m, \forall m \in \Omega_{RSU,} \forall k \in \Omega_{TP}, \forall i \in \Omega_{RSU\cup sensor}, \forall j \in \Omega_{sensor\cup TP} \tag{28}$$

$$f_{mj}^{mk} \leq a_m, \forall m \in \Omega_{RSU,} \forall k \in \Omega_{TP}, \forall j \in \Omega_{sensor\cup TP} \tag{29}$$

$$\sum_{\forall i\in\Omega_{RSU\cup sensor}, \forall j\in\Omega_{sensor}} f_{ij}^{mk} = \sum_{\forall i\in\Omega_{sensor}, \forall j\in\Omega_{sensor\cup TP}} f_{ij}^{mk}, \forall m \in \Omega_{RSU}, \; k \in \Omega_{TP} \tag{30}$$

$$\begin{aligned} &\partial_3 \sum_{\forall i\in\Omega_{RSU,}, \forall j\in\Omega_{sensor}} f_{ij}^{mk} + \partial_2 \sum_{\forall i\in\Omega_{sensor,}, \forall j\in\Omega_{sensor}} f_{ij}^{mk} \\ &+ \partial_1 \sum_{\forall i\in\Omega_{sensor}, \forall j\in\Omega_{TP}} f_{ij}^{mk} < \partial, \forall m \in \Omega_{RSU}, k \in \Omega_{TP} \end{aligned} \tag{31}$$

3 Numerical Analysis

In this section, we evaluate the performance of the DLP problem mentioned in Sect. 4. We first simulate small scenes to verify the correctness and feasibility of the model.

The layout of the network in different scenarios obtained by Gurobi is shown in Figs. 3 and 4, respectively. The solid line shows the communication link between nodes, while the dotted circle shows the coverage. Experiment grid 4 * 20 simulates a road with a length of 200 m and width of 40 m. The parameters are shown in Table 2. We define a generic cost unit (gcu) to assess network costs [7]. The cost of wired connection per meter is 20 gcu, and every BS, RSU, and sensor costs 300 gcu, 80 gcu, and 26 gcu, respectively. Figures 3 and 4 show simulation results in straight road

Table 2. Parameters setting

Parameter	Value
BS_COM	30 km
SN_SEN	31 m
SN_COM	60 m
RSU_R	50 m
$\partial_3, \partial_2, \partial_1, \partial$	0.01, 0.05, 0.09, 1.0(DB/m)

scenario and sharp turn scenario. The coordinates of x-axis and y-axis in Figs. 3 and 4 represent normalized distance, the unit distance shown in these figures represents 10 m in practice.

In this section, we mainly analyze the optimal layout planning of the following two scenarios. Obviously, RSUs, sensors and BSs are selected from candidate points and all

these points are placed in a 200 m * 200 m mesh-graph. Furthermore, RSUs candidate sets and sensors candidate sets are colocated. We take Scenario 1 and Scenario 2 as examples, the total cost of these two scenarios, including entities costs and communication fibers costs, is 1510 gcu and 1511 gcu, respectively. The coverage radius of RSU is larger than the sensor's, so the collaboration between RSUs and sensors decreases the cost of the connection. In addition, we simulate four scenarios in Table 3, although we show two scenarios' layout solutions in this paper, Scenario 3 and Scenario 4 indicate the feasibility and solvability of the proposed architecture. The simulation results indicate that the network architecture is cost-effective and practical in the actual deployment of an Integrated VANET-WSN network.

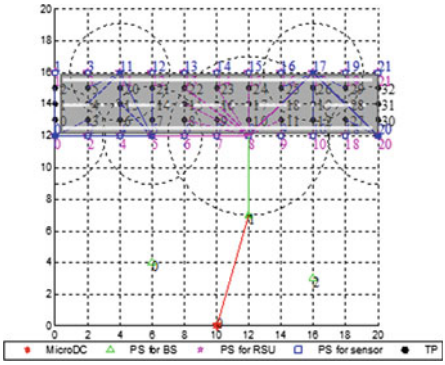


Fig. 3. Layout solution of Scenario 1

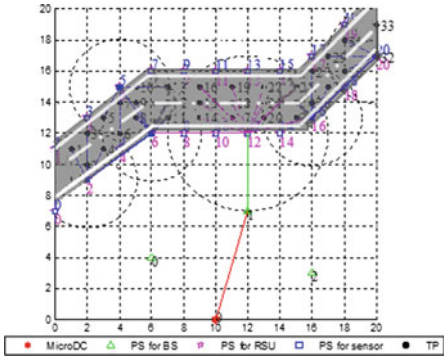


Fig. 4. Layout solution of Scenario 2

Table 3. Scenarios setting and deployment results of different scenarios

Scenario	The number of MicroDC, BSs, RSUs, sensors, TPs	Cost (gcu)
Straight road (Scenario 1)	1, 3, 22, 22, 33	1510
Sharp turn road (Scenario 2)	1, 3, 21, 21, 34	1511
Intersection (Scenario 3)	1, 3, 24, 24, 36	1240
Island road (Scenario 4)	1, 3, 26, 26, 36	1691

4 Conclusion

"Smart" roads based on Integrated VANET-WSN networks could perform well in driving safety, hence the vehicles pass by the dangerous road can receive the road conditions in advance. In this paper, we study the DLP problem for the Integrated VANET-WSN network and get the optimal layout solutions by deploying RSUs and sensors on the roadsides. Simulation results demonstrate the performance in terms of feasibility and scalability of the proposed optimization framework. We find that the VANET-WSN is cost-effective, economical, and practical. The future work is to propose a heuristic algorithm to fast solve the DLP problem in large-scale network scenario of the Integrated VANET-WSN networks.

References

1. Sanguesa, J.A., Naranjo, F., Torres-Sanz, V.: On the study of vehicle density in intelligent transportation systems. Mob. Inf. Syst. **21**(2), 1–13 (2016)
2. Arbi, Z., Driss, O.B., Sbai, M.K.: A multi-agent system for monitoring and regulating road traffic in a smart city. In: 2017 International Conference on Smart, Monitored and Controlled Cities (SM2C). Sfax (2017)
3. Viani, F., Polo, A., Giarola, E., Robol, F.: Performance assessment of a smart road management system for the wireless detection of wildlife road-crossing. In: 2016 IEEE International Smart Cities Conference (ISC2), Trento (2016)
4. Lin, C.C., Chen, P.C., Chang, L.W.: On different-dimensional deployment problems of integrated VANET-sensor networks with QoS considerations. Mob. Netw. Appl. **22**(1), 125–138 (2017)
5. Sadou, M., Bouallouche-Medjkoune, L.: Efficient message delivery in integrated sensor and vehicular networks based on mathematical linear programming. In: Computers & Electrical Engineering (2016)
6. Lin, C.C., Deng, D.J.: Optimal two-lane placement for integrated VANET-sensor networks. IEEE Trans. Ind. Electron. **62**(12), 7883–7891 (2015)
7. Del Castillo, R.A., Garello, N.C., Kuschnir, E.: A multi-objective approach to design all-optical and translucent optical networks considering CapEx and QoT. In: 2012 14th International Conference on Transparent Optical Networks (ICTON), Coventry (2012)

A Survey of Interference Management Technology in Wireless Communication Network

XiaoLin Jiang[1,2(✉)], Ming Diao[2], XiaoJie Chen[1], and YanQiu Du[1]

[1] Heilongjiang University of Science and Technology, Harbin 150022, China
jlynner@163.com
[2] Harbin Engineering University, Harbin 150000, Heilongjiang, China

Abstract. In the next generation mobile communication system, the inter-user interference suppression and elimination in the communication system is an urgent problem to be solved. This paper focuses on the development of wireless network interference management, this paper focuses on the analysis of the status quo and existing problems from the asymptotic capacity analysis technology, multi-cell cooperative transmission technology, MIMO system interference cancellation technology, relay technology, effective spectrum resource detection technology, and to provide better guidance for the study of interference management technology.

Keywords: 5G · Progressive capacity analysis · Relay strategy

1 Introduction

With the growth of communication services, the problem of inter-cell interference in wireless communication network cannot be neglected. Especially for the users at the edge of the cell, because the neighboring cells occupy the same frequency carrier resources, the interference is relatively large, and the distance from the base station Edge of the user service quality is poor, low throughput [1]. The next step 5G communication network will use the same frequency full-duplex technology, that is, in the same spectrum, the communication transceiver both transmit and receive signals at the same time, full-duplex technology to multi-frequency multiplexer and time-division multiplexing Spectrum resource usage restrictions, making the use of spectrum resources more flexible. Through research, the author found that, based on the interference suppression and management problems existing in the existing 4G network, researchers have studied the interference channel asymptotic capacity analysis and interference management techniques used in multi-cellular and large-scale hybrid networks under the condition of full frequency coverage network [2]. For the first time, researchers use this network with asymptotic modeling and analysis capabilities.

This work was supported by the Heilongjiang Province Natural Science Foundations of China (F2015019, F2015017 and E2016061), the Heilongjiang Province Postdoctoral Science Foundations (LBH-Z16054).

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 623–629, 2019.
https://doi.org/10.1007/978-981-13-6264-4_74

On this basis, from the perspective of interference coordination and interference cancellation, researchers study the relay strategies of multi-cell cooperative transmission, interference elimination and interference suppression, and finally achieve effective detection in theoretical and technical research. The following article is on the research of these key research directions to expand the description.

2 Research Hotspot and Status Quo

2.1 Capacity Analysis of Multi-cell Cooperative Communication Network

The modeling and analysis of interference channel and capacity of large-scale interference network is the research basis to solve the problem of interference between edge users in cell, which is of guiding significance to realize interference suppression technology between edge users in cell. But at present, even in simple interference networks, it is very difficult to get accurate capacity area, interference network asymptotic capacity analysis has become an important means.

For the network model with interference such as multi-cell system, capacity analysis under the condition of high SNR becomes an important means to study network capacity because the actual system is a interference-limited system with high SNR. Researchers usually establish multi-user interference broadcast channel model for multi-cell system, and then analyze the system freedom degree for the network with any number of antennas, and the generalized freedom degree for the base station/user with different signal and interference intensity which can better reflect the actual capacity of the system. Researchers have also studied the asymptotic capacity of multi-user and multi-hop interference channels, including the design of multi-hop transmission strategy, the method of intra-hop and inter-hop multi-user interference processing, and so on.

At present, there is no conclusion about the inner bound of freedom (especially the tight inner bound) of multi-cell interference broadcast channel based on partial cooperation of base station. Some articles have proposed to study the influence of partial cooperation of base station on network freedom for multi-cell interference broadcast channel. In the case of complete cooperation, each base station forms a large distributed antenna node, so that the channel model becomes a multi-access or broadcast channel, and its analysis is relatively simple. However, the cost of sharing all data information is relatively high [3], and it is generally only carried out in clusters composed of several adjacent base stations. At this time, the interference between each cluster has an impact on capacity. In the next step, it is suggested to properly model the propagation loss in large-scale networks, and then obtain the asymptotic capacity relationship between network capacity bound and network scale, including power and constant factors, which will be of great significance to this research direction.

2.2 Multi-cell Cooperative Transmission Technology

With the development trend of low cost and low power consumption, the complex signal processing process is gradually moved to the base station. In the research of multi-cell cooperative transmission technology, CoMP transmission technology has been paid more and more attention by academia and industry. It mainly uses the base station cooperative transmission mode to improve the throughput and frequency efficiency of the cell at the edge of the cell, so as to effectively solve the inter-cell interference problem. Base station collaboration is divided into full collaboration and partial collaboration. In the fully cooperative mode, the base station completely interacts with the channel state information and the user data information. The whole system is considered as a distributed antenna virtual cell. This cooperation mode is advantageous for extending the single cell processing method to the multi-cell system. However, complete collaboration requires complete data sharing, which results in a large amount of network overhead and is not suitable for practical systems. Some cooperative methods only require the base station to interact with the channel state information, do not need interactive data, is the focus of the current research. The current multi-cell cooperative beamforming optimization problem is divided into three categories, namely, power optimization, signal-to-noise ratio optimization, and weighting and rate optimization. In [4], the multi-cell model is considered to optimize the power optimization problem into SDP (Semi-definite Programming) form. Some researchers use the Lagrangian dual theory to turn down the power optimization problem into an easy-to-solve dual uplink problem. Researchers use Perron–Frobenius theory to obtain max-min SIN interference (Signal interference noise ratio) The problem of dual problem, weighting and rate optimization is a nonconvex problem, although there is no way to find the optimal solution, some suboptimal algorithms can get better performance. L. Venturino et al. Consider the weight and rate optimization problem under each base station power constraint. First, the structure of the optimal beamforming vector is deduced, and then the KKT condition is used to solve iteratively. The problem of weighting and rate optimization is transformed into weighted and MSE optimization problem by using the mean square error (MSE) and the signal-to-noise ratio. The iterative optimization of the receiving matrix, the weighting matrix, and the sending matrix to obtain the stability of the original optimization problem point. Some researchers have given a method to describe the capacity of interference channel. By setting the user rate ratio, the original problem is transformed into a series of feasibility problems, and two kinds of distributed formulas using successive projection and cyclic projection are given for this feasibility problem Algorithm [5]. In order to solve the problem of power optimization, A. Tolli et al. Introduce multiple variables instead of inter-cell interference items. Each base station performs iterative optimization of local interference and interference between base stations. Some researchers have considered the maximum and minimum signal-to-interference-to-noise ratio (SNR) problem, which satisfies the total power constraint. A two-layer iterative algorithm is proposed by using the inverse relation between the power optimization problem and the SNR optimization problem. In addition, the use of game theory to achieve power optimization, proved that the existence of Nash equilibrium point and the design of a reasonable cost function makes the distributed algorithm and

centralized algorithm to achieve the same performance. All of the above studies assume that the base station can obtain accurate channel state information, and in the FDD (Frequency division duplexing) system, the feedback link is limited bandwidth, the need to design a reasonable and efficient feedback mechanism. It has also been proposed to optimize the feedback bit allocation of local channel and cross channel based on per-cell codebook, and propose a continuous codeword selection algorithm to reduce the quantization error. In order to consider both the finite feedback and the feedback delay, the researcher proposes a feedback bit adjustment strategy based on the received signal strength and the communication link delay. At present, the research on robust multi-cell cooperative transmission with error of channel state information is still in the early stage, and most of it is the bounded error model.

2.3 Interference Elimination Technology for MIMO Systems

In general, information transmitted by a base station of another cell is considered to be interference information with respect to a mobile user. If the transmission data rate of the interference information is low, the mobile user can obtain the data transmission rate by first decoding the interference information and then performing the interference cancellation, but it needs to reduce the data transmission rate of other users. The use of adaptive data rate to ensure the fair transmission of multi-user. In this paper, the reachability of two user interference channels is studied by random coding, and the upper bound of the capacity can be realized in the real system. As a technology to implement interference channel capacity, continuous interference cancellation technology has been extensively studied due to the low complexity of implementation. The interference cancellation strategy of point-to-point single antenna communication with multiple interference sources is also studied. The reach interference area of the jamming channel is realized by using polynomial complexity to solve the interference signal of the source interference signal. This simple decoding strategy is applied to the single antenna multi-user interference channel [6], given the fairness of the data rate between users, the computational polynomial complexity of the distributed iterative method to determine from the adjacent cell data is decoded to eliminate or as a noise treatment. When each user uses multiple codebooks, some decoding can be eliminated at the receiving end, and the interference channel transmission rate can be further improved. When each source using multiple codebooks, the single antenna multi-user interference channel is studied. Considering the user fairness function to determine whether the codebook is decoded or used as an additive noise interference channel capacity implementation method, although the linear speed function is linear increase, the iteration of the algorithm may be ineffective solution. The above research is based on the mobile user, a single antenna or MISO system, which makes the interference receiver of the capacity expression for a constant eigenvalue logarithm, algorithm iteration process can achieve a better iteration of each iteration. For MIMO systems where the mobile subscriber is multi-antenna, the capacity of the receiver is the logarithm of a matrix eigenvalue, and these algorithms cannot be directly extended to the MIMO system.

2.4 Interference Strategy of Interference Suppression

As a candidate technology of LTE-A system, relay technology is regarded as an important means to increase coverage and improve cell edge and cell average throughput. The main function of the relay is to provide service signals for areas with severe shadow fading and coverage, providing coverage of hot spots and indoor coverage. Trunks can provide high data rate coverage, packet mobility, temporary network deployment, cell edge throughput, or provisioning coverage for new areas. The transmission power of the relay is low and the wireless network is connected through the eNB between the core network. This feature allows the relay node to be deployed flexibly and the cost is reduced. However, the introduction of the relay after the system also increased the number of interference sources, so that the whole system of interference problems become more complex. On the one hand, we can reduce inter-cell interference by maximizing the base station to relay and relaying the smaller instantaneous SINR to the user. On the other hand, in order to reduce the multi-cell interference of the relay system, based on the AF (Amplify and forward) relay strategy, the use of the packet transmission control method, that is, the adjacent cell using the transmission time control to achieve [7]. The Adjacent cells in the same time slot to specify a different transmission process, to avoid the adjacent cell relay node at the same time transmission, but the time-division mechanism to reduce the transmission efficiency of the system. The idea of wireless network coding is to use the relay to intercept the interference signal for network coding, and then send the encoded information to each sink, and finally through the edge of the information obtained to eliminate the interference signal. Throughout the communication process, because the wireless network coding compared to the traditional storage and forwarding mechanism using fewer time slots, improve the system's spectral efficiency. In a multi-cell relay system, the relay can receive information of a plurality of base stations. If the relay is network-coded by a plurality of base station signals, it is possible to simultaneously increase the transmission rate of a plurality of users in the cell. Researchers have used relay for multi-source and multi-sink information transmission with backhaul network support. Three kinds of cooperative strategies based on network coding are introduced, namely, finite network coding, linear network coding and grid coding [8]. It has also been combined with a multi-cell broadcasting system using random network coding to develop two effective scheduling algorithms between the relay and the base station to improve the transmission efficiency [9]. Researchers also combine network-based trunking systems with user packets to improve system performance by reducing the power of the interference source [10]. The above methods are to avoid interference between repeaters and users, but in the relay protection system, the time mechanism reduces the multiplexing gain of wireless transmission. The idea of wireless network coding is to use relay to process signals to eliminate interference signals and improve the spectral efficiency of the system. The combination of relay technology and base station to develop two effective scheduling algorithms can improve the performance of the system.

2.5 Study on the Effective Detection of Spectrum Resources

Effective detection of spectrum resources will effectively reduce the interference between users and improve the effective utilization of spectrum, which has a very large application prospect. At present, the wireless static spectrum allocation technology lacks flexibility, which leads to low spectrum utilization and spectrum waste. Dynamic spectrum allocation technology is to analyze the surrounding wireless environment and reuse the free spectrum on the premise of ensuring the normal communication of authorized users, so as to use the limited spectrum resources more effectively and alleviate the contradiction between wireless communication needs and spectrum resources. Therefore, dynamic spectrum allocation technology has a very broad application prospect in 4G wireless communication field and has become an effective way of interference management technology and a research hotspot. The spectrum detection technology has also been extensively studied at home and abroad, among which the algorithm performance based on energy detection is the most widely studied. In combination with some new technologies, some breakthroughs have been made. Some researchers have proposed to combine stochastic resonance with energy detection to improve spectral perception performance of energy detection effectively by using stochastic resonance technology. Some researchers analyzed the improvement of noise wall based on stochastic resonance energy detection for energy detection. Performance analysis of output SNR of bistable stochastic resonance system; Optimal binary quantization detector and asymptotically optimal array threshold detector for generalized stochastic resonance system; In addition to the classical spectral perception algorithm and the improvement of the classical spectral perception algorithm, some new spectral perception improvement algorithms have also been studied in many literatures. For example, the wavelet-based spectrum sensing method, filter bank spectrum estimation, and so on.

3 Summary

In the wireless communication system, the interference suppression and elimination is an urgent problem to be solved. At present, on the interference channel capacity analysis and large-scale interference network capacity asymptotic analysis is a simple network or simple protocol for a rough analysis, cannot fully reflect the actual interference network capacity area. CoMP technology can improve the system rate by interacting with channel state information and base stations, but most of the research is aimed at CoMP strategy under centralized and ideal channel. On the other hand, the current interference cancellation technology only considers MISO Network; at the same time, relay technology in order to avoid inter-channel interference using time-division strategy to transmit, which reduces the multi-cell communication transmission rate. Therefore, the asymptotic analysis, distributed and robust CoMP technology for interference channel capacity, the low complexity of MIMO system to achieve interference cancellation technology and enhance the edge of the user transmission rate of the relay strategy constitutes the next generation of mobile communication system interference management The key technology to carry out this research is of great significance.

References

1. 3GPP: Huawei: Soft frequency reuse scheme for UTRAN LTE. R1-050507, May 2005
2. Dahrouj, H., Yu, W.: Coordinated Beamforming for the multicell multi-antenna wireless system. IEEE Trans. Wirel. Commun. **9**, 1748–1759 (2010)
3. Cadambe, V.R., Jafar, S.A.: Interference alignment and the degrees of freedom of wireless X networks. IEEE Trans. Inform. Theor. **9**, 3893–3908 (2016)
4. Dahrouj, H., Yu, W.: Coordinated Beamforming for the multicell multi-antenna wireless system. IEEE Trans. Wirel. Commun. **9**, 1748–1759 (2015)
5. Qiu, J.M., Zhang, R., Luo, Z.Q., Cui, S.G.: Optimal Distributed Beamforming for MISO interference channels. IEEE Trans. Signal Process. **59**, 5638–5643 (2011)
6. Tajer, A., Prasad, N., Wang, X.: Fair Rate Adaptation in Multiuser Interference Channels. IEEE ISIT, Texas, USA (2010)
7. Gong, C., Tajer, A., Wang, X.: Interference channel with constrained partial group decoding. IEEE Trans. Commun. **59**(11), 3.59–3.71 (2011)
8. Du, J., Xiao, M., Skoglund, M.: Cooperative strategies for relay-aided multi-cell wireless networks with backhaul. In: Proceedings of IEEE Information Theory Workshop (ITW) (2010)
9. Özbek, B., Ruyet, D.L.: Adaptive limited feedback links for cooperative multi-antenna multicell networks. Eurasip J. Wirel. Commun. Netw. **1**, 1–11 (2014)
10. Mane, P.R.: Analysis and performance enhancement of network coding for wireless networks. Manipal (2014)

A Novel Waveform for Radar and Communication Integration in the Satellite System

Liang Xu, Yicheng Jiang(✉), and Zhaofa Wang

Harbin Institute of Technology, 92 West Dazhi Street, Nangang District, Harbin 150001, China
Jiangyc@hit.edu.cn

Abstract. The integration of radar and communication in the satellite system can realise the functions of target detection and wireless communication operations simultaneously. The key technique to achieve the integration is to design an integrated transmitted waveform. A novel transmitted waveform for a satellite system is designed in this paper. The waveform is based on the triangular frequency modulated continuous wave (FMCW) and the quadrature phase shift keying (QPSK). The ambiguity function (AF) and the bit error rate (BER) of the waveform are derived to demonstrate the feasibility. The simulated results validate the performance of the integrated waveform in this paper.

Keywords: Radar-communication integration · Waveform design · Ambiguity function · Bit error rate

1 Introduction

In recent years, the radar-communication integration has received more and more attention for its good performance [1]. The integration can reduce the number of onboard devices, decrease the risk of electromagnetic interference. Moreover, an integrated system costs much less than two independent systems [2].

In general, the existing radar-communication integration methods can mainly be divided into two categories. One is that two functions share the same hardware system, and their corresponding signals are transmitted separately to perform its function. The radar and communication signals can be transmitted in different space domain [1], time slots [2] or frequency bands [3]. In these methods, the receivers have to separate the two signals first, which increases the complexity. The other category is that the two functions share the hardware system, and use the same transmitted signal. The key technique of this kind of integration methods is the design of the integrated waveform. A kind of integrated waveform based on the Orthogonal Frequency Division Multiplexing (OFDM) is proposed, which can be used for target detection and data transmission directly [4, 5]. However, the OFDM waveform is sensitive to the Doppler shift and its high peak-to-mean envelope power ratio restricts the detection range [6]. Another kind of integrated waveform is based on the linear frequency modulated (LFM) waveform. In [7], the integrated waveform is designed by modulating the data

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 630–637, 2019.
https://doi.org/10.1007/978-981-13-6264-4_75

onto the LFM waveform by the minimum shift keying (MSK) modulation scheme, where the LFM waveform is used as the carrier to ensure the radar performance of the integrated waveform, and the MSK method is used to encode data to support the data transmission. In [8, 9], the amplitude modulation (AM) and the binary phase shift keying (BPSK) modulation schemes are, respectively, employed to add data onto the LFM signal, and the results show their good performance in both radar and communication. In addition, the existing radar-communication integration methods are hardly based on a communication system, especially a satellite communication system.

In order to integrate the radar and communication functions into a geosynchronous satellite system, we design a novel transmitted waveform which is based on the triangular FMCW signal and the QPSK modulation scheme, where the triangular FMCW signal is used as the carrier wave to perform the radar function, and the QPSK modulation scheme, which is widely used in the geosynchronous satellite communication system, is utilised to modulate the data. To analyse the feasibility of the integrated waveform, the ambiguity function (AF) and bit error rate (BER) of the waveform are investigated, where the AF is used to test the target detection performance, and the BER is used to examine the communication capacity. The simulations are run to demonstrate the performance of the integrated waveform.

The remainder of this paper is organised as follows. In Sect. 2, the system model is established and the novel waveform is proposed. In Sect. 3, the AF of the signal is derived and the resolutions of the time delay and the Doppler shift are analysed. The simulation results and analysis are provided in Sect. 4. Conclusions are given in Sect. 5.

2 System Model and Waveform Design

2.1 System Model

The geometry of the geosynchronous satellite system is shown in Fig. 1, where L denotes the length of baseline, R_T and R_R are the distances from target to transmitter and receiver, respectively, θ_R denotes the look angle of the receiver called the elevation angle, β is the bistatic angle, and ϕ represents the angle between the target velocity vector V and the bistatic bisector of β.

According to the geometry of the system, the satellite transmits the integrated signal into the illumination area. The receiver receives the target echo and the direct signal simultaneously, where the two kinds of signal can be used for target detection and data transmission, respectively. It can be seen that the multifunctional satellite system is the combination of a traditional satellite communication system and a bistatic radar system.

From the analysis above, the traditional radar or communication signals cannot be used to achieve the two functions simultaneously. To fit the system, an integrated waveform needs to be designed.

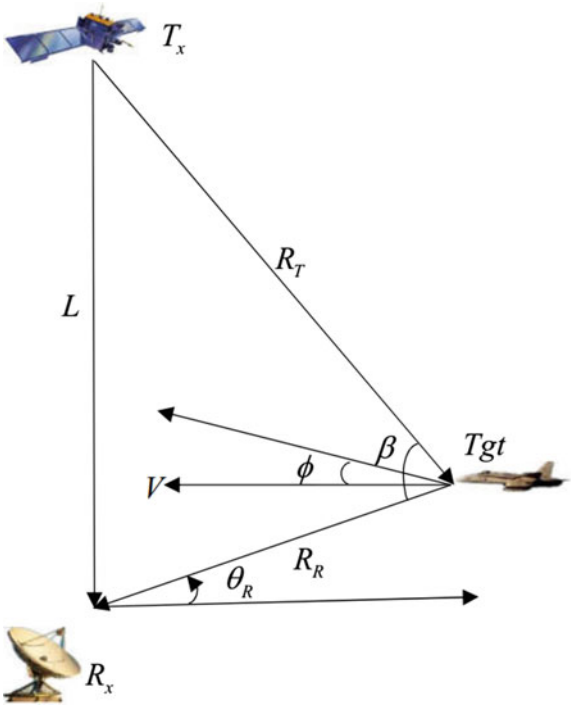


Fig. 1. Geometry of the system

2.2 Waveform Design

An efficient way to design the shared waveform is to add data onto the radar signal. In this paper, the integrated waveform is based on the triangular FMCW signal and the QPSK modulation scheme, where the triangular FMCW signal is chosen for its advantage of eliminating the range and Doppler coupling.

A period of the triangular FMCW signal is consists of an 'up-chirp' and a 'down-chirp', it can be expressed as

$$\begin{aligned} f(t) = & \, rect\left(\frac{t}{T/2}\right) e^{j(2\pi f_0 t + \pi \mu t^2)} \\ & + rect\left(\frac{t - T/2}{T/2}\right) e^{j[2\pi (f_0 + 2B)t - \pi \mu t^2]} \end{aligned} \tag{1}$$

where

$$rect(t) = \begin{cases} 1, & 0 \le t < 1 \\ 0, & else \end{cases},$$

T represents the duration of a period, f_0 is the carrier frequency, B is bandwidth, and $\mu = 2B/T$ denotes the frequency modulate slope.

The kth symbol of QPSK-FMCW signal can be written as

$$s_k(t) = rect\left(\frac{t-(k-1)T_{sym}}{T_{sym}}\right)f(t)e^{j\theta_k} \tag{2}$$

where T_{sym} is the symbol width and $T_{sym} = T/N$, N is the symbol number in a period, and the data rate $R_b = \frac{2N}{T}$ bit/s, θ_k denotes the quaternary communication data.

Then a period of QPSK-FMCW signal is given by

$$s(t) = s^+(t) + s^-(t) \tag{3}$$

where

$$s^+(t) = \sum_{k=1}^{N/2} rect(\frac{t-(k-1)T_{sym}}{T_{sym}})e^{j(2\pi f_0 t + \pi\mu t^2 + \theta_k)} \tag{4}$$

$$s^-(t) = \sum_{k=N/2+1}^{N} rect(\frac{t-(k-1)T_{sym}}{T_{sym}})e^{j(2\pi(f_0+2B)t - \pi\mu t^2 + \theta_k)} \tag{5}$$

3 The Ambiguity Function (AF) Analysis

As a significant tool, the AF is usually used to analyse the performance of a radar signal [10]. In order to examine the radar function of the integrated waveform in the satellite system, the AF analysis is used in this section. As the integrated waveform employs the QPSK modulation, and the satellite system works as a bistatic radar system. In this section, we derive the AFs of the integrated signal.

For a transmitted signal $f(t)$, its ambiguity function is defined as follows:

$$|\chi(\tau, f_d)|^2 = \left|\int_{-\infty}^{+\infty} \tilde{f}(t)\tilde{f}^*(t-\tau)e^{j2\pi f_d t}dt\right|^2 \tag{6}$$

where $\tilde{f}(t)$ denotes the complex envelope of the transmitted signal, τ is the time delay, and f_d is the Doppler shift.

According to (3–5), the complex envelope of the QPSK-FMCW signal is

$$\begin{aligned}\tilde{s}(t) &= \tilde{s}^+(t) + \tilde{s}^-(t) \\ &= \sum_{k=1}^{N/2} rect(\frac{t-(k-1)T_{sym}}{T_{sym}})e^{j(\pi\mu t^2 + \theta_k)} \\ &\quad + \sum_{k=N/2+1}^{N} rect(\frac{t-(k-1)T_{sym}}{T_{sym}})e^{j(-\pi\mu t^2 + \theta_k)}\end{aligned} \tag{7}$$

Let τ be around 0, we have $T_{sym} - \tau \approx T_{sym}$, The AF of QPSK-FMCW signal can be expressed as

$$\begin{aligned} |\chi_s(\tau, 0)|^2 &\approx \left| \frac{1}{2T} \left[NT_{sym} \sin c\left(\mu \tau NT_{sym}/2\right) \right.\right. \\ &\quad \left.\left. + NT_{sym} \sin c\left(-\mu \tau NT_{sym}/2\right) \right] \right|^2 \\ &= \left| \sin c(B\tau) \right|^2 \end{aligned} \tag{8}$$

4 Simulation and Analysis

In this section, the performance of the integrated waveform, including the radar function and the communication function, is simulated and analysed. The AF of the integrated waveform is simulated to analyse the target detection performance. And the communication performance is validated by comparing the BER of the QPSK-FMCW signal and the QPSK signal.

4.1 Simulation of Radar Function

As analysed in Sect. 3, the AF of the QPSK-FMCW signal used in the satellite system is affected by the QPSK modulation and the system geometry. Therefore, based on the AF of the FMCW signal, the influence of the QPSK modulation and the system geometry are tested, respectively.

In order to examine the influence of the QPSK modulation, the AFs of the FMCW signal and the QPSK-FMCW signal are compared. For the AF, a thumbtack shape means the narrow mainlobe and low-level sidelobes, which leads to a better performance of the signal in target detection. Thus, the shape, the width of the mainlobe and the level of the sidelobes of the two AFs are compared, respectively.

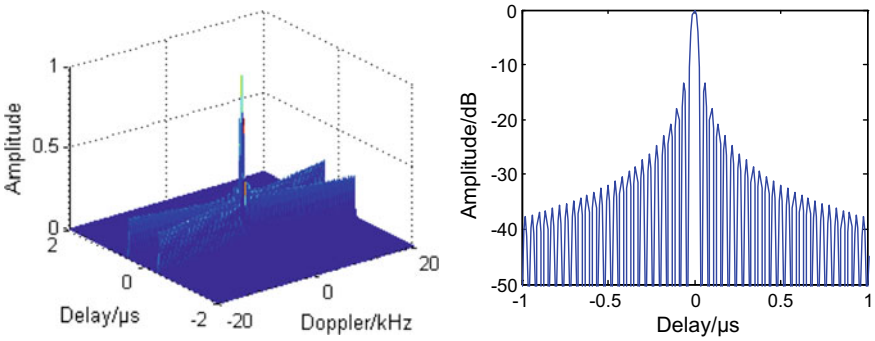
The parameters of the QPSK-FMCW signal are as follows: the carrier frequency $f_0 = 1.5$ GHz, the duration $T = 1$ ms, the bandwidth $B = 25$ MHz, and the data rate $R_b = 2$ Mb/s.

Figures 2 and 3 show the AFs of the FMCW signal and the integrated signal, respectively. Comparing the simulation results in Figs. 2 and 3, it can be seen that: (1) the AF of the QPSK-FMCW signal is closer to a thumbtack shape than that of the FMCW signal, (2) the data modulation only affects the sidelobes of zero Doppler slice.

4.2 Simulation of Communication Function

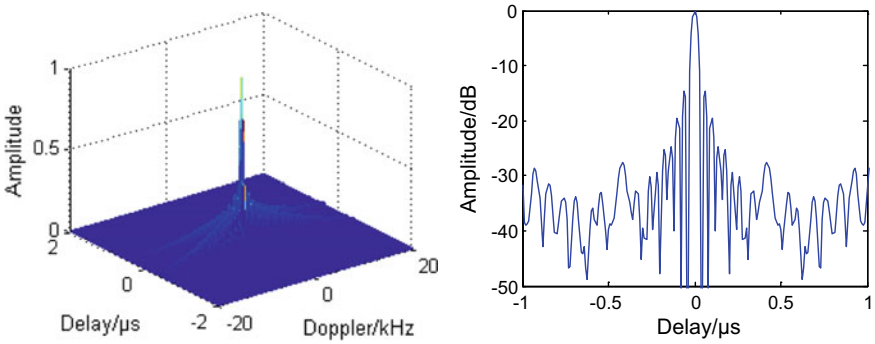
The BER is a widely-used metric to evaluate the reliability of communication [2], and it depends on the signal-to-noise ratio (SNR) and the parameters of the signal. In order to verify the communication performance of the QPSK-FMCW signal, the BER results of the QPSK-FMCW signal and the QPSK signal are compared.

The demodulation of the QPSK-FMCW signal is the same as that of the QPSK signal, the frequency modulation on the carrier does not affect the data transmission.



(a) The AF (b) Zero Doppler slice

Fig. 2. The AF of the FMCW signal



(a) The AF (b) Zero Doppler slice

Fig. 3. The AF of the integrated signal

Thus the theoretical results of the BER probability of the two signals are identical. It is assumed that the parameters of the QPSK-FMCW signal are the same as Sect. 4.1. The carrier frequency and the data rates of the QPSK signal are the same with the QPSK-FMCW signal. The channel is modelled as an additive white Gaussian noise (AWGN) channel. Figure 4 shows the comparison results of the BER performance of the two signals.

As shown in Fig. 4, the BER performance of the QPSK-FMCW signal is the same as the QPSK signal, which indicates that using othe FMCW as carrier wave will not affect the BER performance. In other words, the integrated signal can be used to implement the communication function.

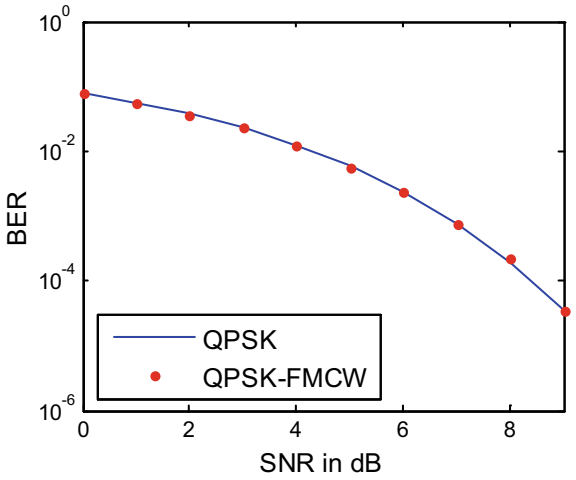


Fig. 4. The BER performance of the two signals

5 Conclusions

In this paper, a novel transmitted waveform is designed to achieve the radar and communication functions in a satellite system. The AF of the integrated waveform is derived and the influence of the geometry of the system is analysed. Simulation results show that the integrated waveform has a good performance in the target detection and communication.

Acknowledgements. This paper is supported by the China Scholarship Council and the National Natural Science Foundation of China (Grant No. G61271342).

References

1. Tavik, G.C., Hilterbrick, C.L., Evins, J.B., et al.: The advanced multifunction RF concept. IEEE Trans. Microw. Theor. Techn. Mar. **53**(3), 1009–1020 (2005)
2. Han, L., Wu, K.: Multifunctional transceiver for future intelligent transportation systems. IEEE Trans. Microw. Theor. Techn. **59**(7), 1879–1892 (2011)
3. Winkler, V., Detlefsen, J.: Automotive 24 GHz pulse radar extended by a DQPSK communication channel. In: Radar Conference, 2007. EuRAD 2007. European. IEEE, 2007, pp. 138–141 (2007)
4. Gu, J.F., Moghaddasi, J., Wu, K.: Delay and Doppler shift estimation for OFDM-based radar-radio (RadCom) system. In: 2015 IEEE International Wireless Symposium (IWS), pp. 1–4. IEEE (2015)
5. Hu, L., Du, Z., Xue, G.: Radar-communication integration based on OFDM signal. In: 2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), pp. 442–445. IEEE (2014)

6. Franken, G.E.A., Nikookar, H., Van Genderen, P.: Doppler tolerance of OFDM-coded radar signals. In: Proceedings of the 3rd European Radar Conference, pp. 13–15 (2006)
7. Chen, X., Wang, X., Xu, S., et al.: A novel radar waveform compatible with communication. In: IEEE Conference on Computational Problem-Solving (ICCP), Chengdu, China, pp. 177–181 (2011)
8. Barrenechea, P., Elferink, F., Janssen, J.: FMCW radar with broadband communication capability. In: Radar Conference, 2007. EuRAD 2007. European, pp. 130–133 (2007)
9. Zhao, Z., Jiang, D.: A novel integrated radar and communication waveform based on LFM signal. In: 2015 5th International Conference on Electronics Information and Emergency Communication (ICEIEC), pp. 219–223. IEEE (2015)
10. Tsao, T., Slamani, M., Varshney, P., et al.: Ambiguity function for a bistatic radar. IEEE Trans. Aerosp. Electron. Syst. **33**(3), 1041–1051 (1997)

Construction of Microwave Network Topology Based on Disjoint Path Algorithm

Li Shang[1(✉)] and Zhou Ji[2(✉)]

[1] Institute of Computer Application, China Academy of Engineering Physics, Mianyang, China
41540356@qq.com

[2] China Electronic Technology Group Corporation Twenty-Eighth Research Institute, Nanjing, China
zhouji@hitee.net

Abstract. In order to construct a battlefield microwave network topology with high reliability, low power, and low interference, a topology construction algorithm is proposed in this paper. While satisfying the requirements for network-connectivity and network-disjoint-paths, the algorithm uses the reliability of two nodes and the whole network as the evaluation index, so that transmission power is greatly reduced. Simulation results show that this algorithm can effectively reduce both the node power and redundant network topology compared with k-connected network topology construction algorithm.

Keywords: Microwave network topology · Disjoint path · Network reliability · Low power consumption

1 Preface

As the link of the information warfare command system, the tactical communication network is designed to realize the interconnection of the command units at all levels in the battlefield environment. According to the actual demand of the battlefield, the network topology should have certain reliability; at the same time, the transmission power and the interference [1] between network nodes should be reduced as much as possible. Building a reliable microwave network topology is the core part of tactical communication network.

Evaluating the reliability of the network is the basis of the network topology, hence different reliability evaluation standards correspond to different network construction methods. Network reliability includes effectiveness, survivability, etc. However, the existing methods evaluate reliability from the angle of network connectivity. They fail to measure the destruction resistance of nodes or links in the network topology. Thus their performance indexes can only reflect the network destruction resistance ability from the side [2].

Currently, the state-of-the-art network topology construction methods mainly include the algorithm based on k-connectivity and the algorithm [3] based on connectivity constraints. Most algorithms are based on directivity, connectivity graph and node-based-degree. Among these algorithms, XY Li et al. proposed the YAO algorithm

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 638–645, 2019.
https://doi.org/10.1007/978-981-13-6264-4_76

by first considering how to associate the transmission range with the node number in the fixed area, so that when all nodes have the same transmission range, the obtained network can maintain the K fault node [4] in its neighbor with a high probability. Besides, the MST algorithm is based on the minimum generation. Actually, it is a tree topology control algorithm used in wireless multi-hop networks. In MST, each node independently builds its local minimum spanning tree and only preserves the node on the neighbor hop in the last topology of the distance [5]. CBTC algorithm requires that the node is transmitted with the minimum power to ensure that at each angle a node can be implemented with minimal power [6]; and the K-NEIGH algorithm adopts the principle that the number of physical neighbors around each node is equal to or slightly lower than a specific value (K), which allows for non-boundary physical nodes to produce bounded interferences [7]. The aforementioned methods are controlled and optimized according to the conditions of 1-connectivity, k-connectivity or probability connectivity. But they do not take into account the performance indexes such as the resistance and reliability among the nodes in the network. What's more, these connectivity constraints are rough indicators for network performance and it is difficult for them to evaluate the damage resistance and reliability between different links or nodes in different forms of networks.

In summary, the above algorithms are not very operable for actual requirements of tactical communication network. To establish a novel tactical communication network topology, we propose a method which considers the reliability constraints of the whole network and the reliability requirements between different links or nodes. By invoking the graph theory, a topology construction algorithm based on disjoint path number and the minimum node transmission power is proposed, in which way a network topology inbuilt that not only meets the required reliability and invulnerability requirements but also minimizes power and interference.

2 Reliability Evaluation of Microwave Network

Most of the existing reliability evaluation algorithms take into account the reliability of the topology of the entire microwave network, but they ignore the effect of one or more nodes or link failures on the entire network. Alternatively, this paper will propose a network reliability evaluation standard based on the number of disjoint paths and minimum transmission power. According to the Menger theorem, disjoint paths are two paths with no common nodes in the network topology diagram G (V, E), which are also named as the intersecting path [8]. Via the number of disjoint paths, the reliability evaluation index in this paper realizes the reliability requirements of the whole topology and the internal nodes, while minimizing the interference between different network nodes by power control. The specific evaluation indexes are as follows:

$$\begin{cases} \min \cdot \max\{TP(v_{pi}), v_{pi} \in V(G)\} \\ subject\ \ to: \\ E_{rel}(v_{pi}, v_{pj}) \geq C_{pi,pj} \qquad (pi, pj \in V, pi \neq pj) \\ E_{rel}(G) \geq C_G \end{cases} \tag{1}$$

$$\begin{cases} E_{rel}(v_{pi}, v_{pj}) = \dfrac{N_{pi,pj}}{N_{pi,pj}(\max)} & (pi, pj \in V, pi \neq pj) \\ E_{rel}(G) = \dfrac{\sum\limits_{\substack{pi,pj \in V \\ pi \neq pj}} N_{pi,pj}}{\sum\limits_{\substack{pi,pj \in V \\ pi \neq pj}} N_{pi,pj}(\max)} \end{cases} \tag{2}$$

In the above equations, $TP(v_{pi})$ refers to the upper limit of the transmitting power of the nodes which satisfy the network topology.

$C_{pi,pj}$ and C_G respectively represent the network reliability index of nodes p_i, p_j and the whole network, under the condition that the network topology is fully connected.

$E_{rel}(v_i, v_j)$ and $E_{rel}(G)$ stand for the reliability index of the two nodes in the network topology and that of the whole network, respectively;

$N_{pi,pj}$ refers to the number of disjoint paths in the network topology.

$N_{pi,pj}(max)$ refers to the number of disjoint paths between two nodes when the network is fully connected.

3 Topology Construction of Microwave Network

3.1 Network Topology Construction of Connectivity Based K-Connectivity Algorithm

It is assumed that the microwave network topology is made up of n nodes. Additionally, the connection graph of the network is represented by G(V,E).

Definition 1 The connection matrix L of the microwave network topology is n-order with the mathematical expression:

$$L = \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1n} \\ l_{21} & l_{22} & \cdots & l_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix} \tag{3}$$

where l_{ij} represents the connection of nodes in the network. l_{ij} = 1indicates that node i is linked with the node j; l_{ij} = 0 means the node i and node j are non-connected. Since a microwave station cannot be self-connected,l_{ij}= 0 when $i = j$. In the algorithm, we will use two connection matrices, i.e., all-connection matrix and current-connection matrix, which are represented by L_{all} and $L_{cur,}$ respectively. In the all-connection matrix, l_{ij} = 1 when i is not equal to j.

Definition 2 The weight matrix D of the microwave network topology is a n-order matrix, expressed as

$$D = \begin{bmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{1n} \\ d_{21} & d_{22} & d_{23} & \cdots & d_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ d_{n1} & d_{n2} & d_{n3} & \cdots & d_{nn} \end{bmatrix} \tag{4}$$

The value of d_{ij} represents the length of the microwave link between node i and node j. When two nodes are not connected, (e.g., $i = j$),d_{ij} = INF,where INF is an infinite number to indicate that connection does not exist. The algorithm needs to use two weight matrices, i.e., D_1 and D_2, in which D_1 is a matrix representing the distance between nodes in the topology aimed for finding the shortest path and D_2 represents the total microwave link which is used to obtain the shortest link.

The specific steps based on the k-connectivity algorithm are as follows:

Step 1: Initialize the current-connection matrix L_{cur} and all-connection matrix matrix L_{all}; initialize distance weight matrix D_1 and link length weight matrix D_2.

Step 2: Find the total minimum value of the link set in the weight matrix D_1, and record it as d_{ij}, define $S = i$, $T = j$.

Step 3: If d_{ST} is equal to INF, it will continue; if not, the microwave link does not exist, output L_{cur} and end.

Step 4: According to the shortest path algorithm, calculate the connectivity N_0 in the matrix L_{cur} between and S and T, which are obtained in Step 2.

Step 5: Judge whether the connectivity degree N_0 obtained in step 4 reaches the expected requirement, if it is true, delete the microwave link, then execute $d_{TS} = d_{ST} =$ INF and return to step 2; if false, the missing path of the all-connection matrix L_{all} is found in the weight matrixD_2 to represent the current topology matrix L_{cur}, and then $d_{TS} = d_{ST} =$ INF, also go back to step 2.

The specific execution process of the algorithm is shown in Fig. 1.

3.2 Network Topology Construction Based on Disjoint Path Algorithm

This paper further improves the network topology construction algorithm based on the node power minimization algorithm in [9]. The optimization goal of this algorithm is to reduce the transmission power as much as possible while satisfying the network connectivity and disjoint path requirements. If the influence of the terrain is not considered, the emission power determines the maximum overlapping radius of the node. Hence, the goal is converted into minimizing the maximum emission radius of the node.

The core idea of this paper is as follows: first, the coverage range of nodes is determined according to their transmitting power, and then a fully connected network is constructed according to the coverage of different nodes. After this, the number of disjoint paths is calculated according to the obtained connected graph. If the number of disjoint paths is larger than the requirement, then the node with the largest transmission power should be deleted. If the number of disjoint paths is not enough, a microwave relay station algorithm is needed to construct adaptive microwave links. The algorithm uses the maximum flow minimum cut algorithm and the minimum spanning tree algorithm in graph theory to resolve the number of disjoint paths and finds the shortest

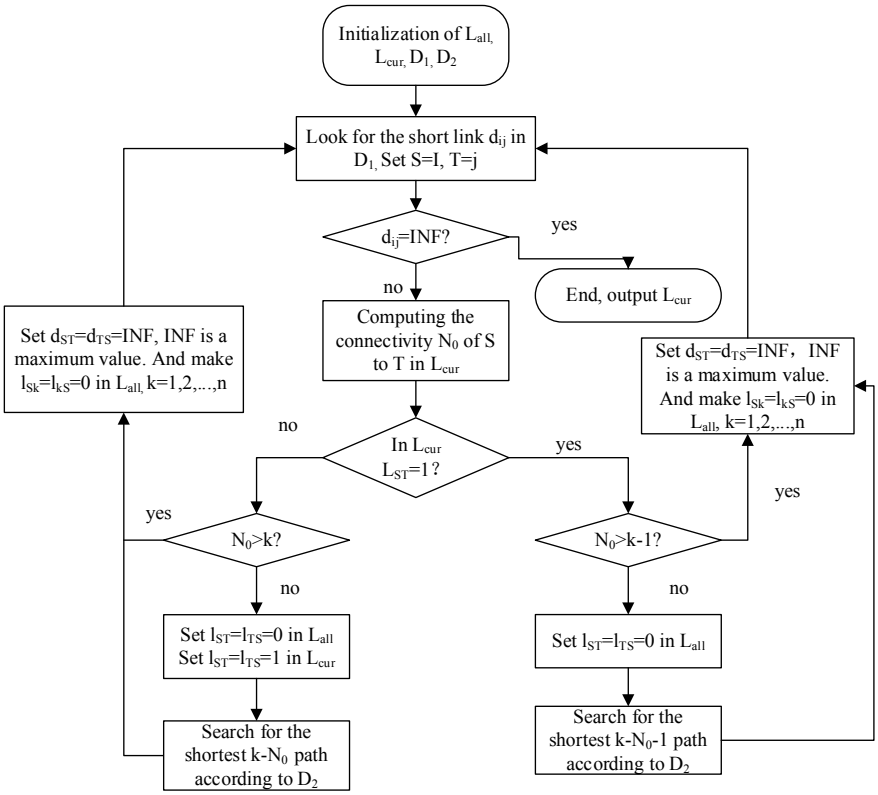


Fig. 1. k-connectivity algorithm flow

path by the minimum spanning tree algorithm. Therefore, the resulted network topology satisfies both the shortest and the minimum transmission power while ensuring connectivity. The algorithm flow is demonstrated in Fig. 2.

The explicit steps of the algorithm are as follows:

Step 1: according to the transmission power of each node in the network, determine the radius of the launch, and build a basic connection graph, which is called $G(V, E)$.

Step 2: split the basic connection graph $G(V, E)$:

(1) Split each node v_i in the network connection diagram $G(V, E)$ into v_{i1} and v_{i2}, and then assumes that there is a link between the two nodes with the capacity of 1.
(2) Each link v_iv_j in the network connection diagram $G(V, E)$ is split into two links including $v_{i1}v_{j1}$ and $v_{i2}v_{j2}$, and the capacity of the link is 1.

Step 3: mark the source node v_s with ds = ∞ and ls = −2 let S be the set of nodes that have been traversed(S = $\{v_s\}$), define the connection matrix by E, where the initial value of E is zero matrix, with i = s。

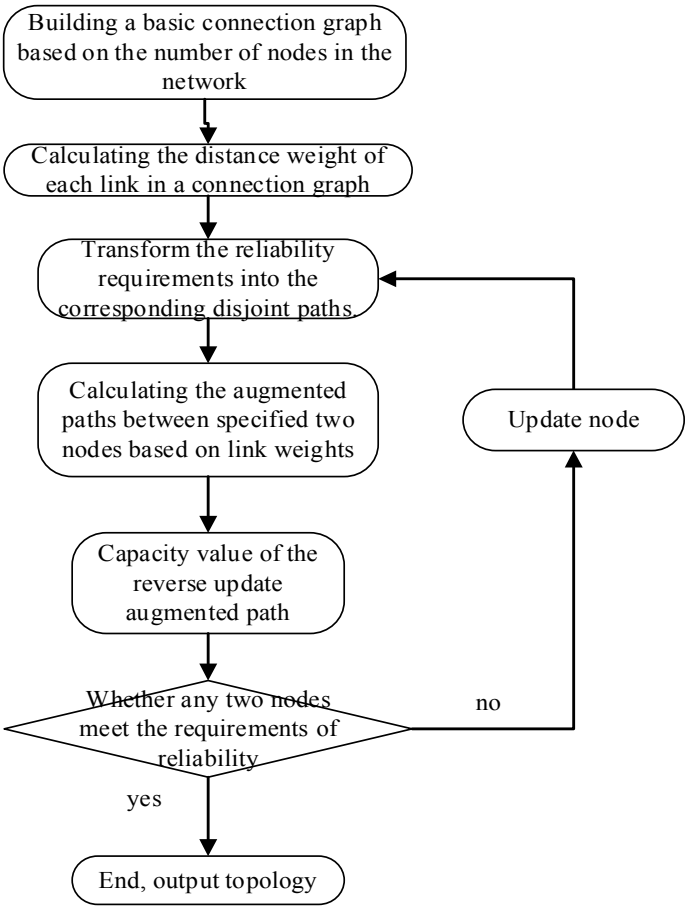


Fig. 2. Algorithm for minimizing node power and disjoint paths

Step 4: mark the destination node v_i with $v_i = (l_i, d_i)$, use the matrix $[S,S']$to judge the connectivity between the source node and the destination node. If $[S,S']$ is a zero matrix, the algorithm terminates, otherwise get the minimum link via $\min(v_i, v_j)$。

Step 5: Set $d_j = \min\{d_i, c_{ij}(f_k)\}$ and judge whether j is equal to t. If they are equal, go to the next step; if not, return to Step 4.

Step 6: Get the reverse link $c_{ij}(f_k) = c_{ij}(f_k) + f_k$ and the positive link $c_{ij}(f_k) = c_{ij}(f_k) + f_k$ by augmenting the path and recording the feasible path for the target node and the nodes along the reverse path.

Judge whether k satisfies the requirement for reliability and disjoint path number. If it satisfies, end the algorithm, otherwise, go back to the step and continue the search.

Step 7: Record the above result, and the paths under the constraints of reliability and destruction of any two nodes are repeated, and the above steps are repeated until the network topology satisfies requirements for the specified disjoint path number and reliability.

4 Simulation Analysis

In this paper, the simulation analysis is made on two aspects, i.e., the k-connectivity-based topology construction algorithm in Sect. 1 (algorithm 1) and the topology construction algorithm in Sect. 2 which is based on the minimum node power and disjoint path number (algorithm 2). Take the electronic map in Fig. 3 as an example. 10 nodes are selected randomly in the map to set up the transmitting power. With the same positions and the same parameters of 10 nodes, the above two methods are adopted, respectively. The results are shown in Figs. 3 and 4.

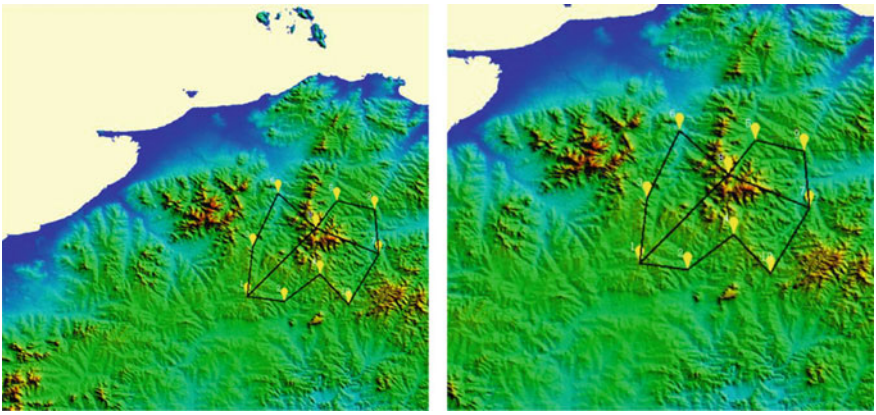


Fig. 3. Simulation results of algorithm 1

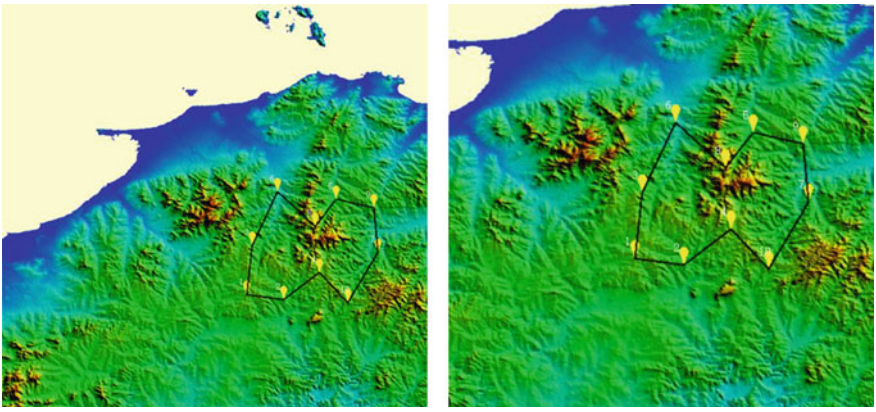


Fig. 4. Simulation results of algorithm 2

By comparing the simulation results, we can see algorithm 1 and 2 provide similar but different typology network designs under the same condition. On the premise of satisfying the same disjoint path number, the link number of the topology control

algorithm based on the node power minimization and the disjoint path (in Fig. 4) is less than that of the k-connected network topology construction algorithm (in Fig. 3), while the reliability is not reduced. Hence, the superiority of the proposed algorithm (algorithm 2) is validated.

References

1. Deniz, F., Bagci, H., Korpeoglu, I.: An adaptive, energy-aware and distributed fault-tolerant topology-control algorithm for heterogeneous wireless sensor networks. Ad Hoc Netw. **44**(C), 104–117 (2016)
2. Gui, J., Zhou, K.: Flexible adjustments between energy and capacity for topology control in heterogeneous wireless multi-hop networks. J. Netw. Syst. Manag. **24**(4), 789–812 (2016)
3. Huang, J., Duan, Q., Xing, C.C., et al.: Topology Control for Building a Large-Scale and Energy-Efficient Internet of Things[J]. IEEE Wirel. Commun. **24**(1), 67–73 (2017)
4. Li, X.Y., Wan, P.J., Wang, Y., et al.: Fault tolerant deployment and topology control in wireless networks. Wirel. Commun. Mobile Comput. **4**(1), 117–128 (2004)
5. Li, N., Hou, J.C., Sha, L.: Design and analysis of an MST-based topology control algorithm. IEEE Trans. Wirel. Commun. **4**(3), 1195–1206 (2005)
6. Li, L., Joseph, Y.H., Paramvir, B., et al.: A cone-based distributed topology-control algorithm for wireless multi-hop networks. IEEE Trans. Netw. **13**(1), 147–159 (2005)
7. Blough, D.M., Leoncini, M., Resta, G., et al.: The k-neighbors approach to interference bounded and symmetric topology control in Ad Hoc networks. IEEE Trans. Mobile Comput. **5**(9), 1267–1282 (2006)
8. Ruiying, L., Rui, K.: Summary of research on network reliability evaluation. Reliab. Eng. **7**(3), 104–108 (2008)
9. BaoXuecai. Resource allocation optimization for wireless Mesh networks with reliability constraints. Harbin Institute of Technology, pp. 13–16 (2013)

A Resource Scheduling Algorithm Based on Service Buffer for LTE-R Network

Yu Xing(✉), Gang Chuai, Weidong Gao, and Qian Liu

Key Laboratory of Universal Wireless Communications Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China
{xyu,chuai,gaoweidong,LiuQian}@bupt.edu.cn

Abstract. With the rapid development of high-speed railway, the high-speed railway network has gradually increased its requirements for communication quality. There are still some shortcomings in the current network at present, such as inflexible resource scheduling, poor reliability, and inconvenient operation and maintenance. The system requires high reliability and real-time CBTC (communication-based train control system) services; the demand for throughput of the necessary video services is very high. Therefore, it is necessary to design a radio resource scheduling strategy that is specifically proposed for the system in order to more efficiently use radio resources to improve system performance. According to the predicted railway communication situation, a multiservice bearer model based on a private network is used. In this mode, the RAD [1] algorithm cannot meet all the requirements of system performance. This paper proposes an improved RAD resource scheduling algorithm based on service cache. Simulation results show that compared with RAD algorithm, our new improved algorithm achieves better packet call throughput and delay.

1 Introduction

The private network of high-speed railway communication is important to ensure the safe operation of train, as it demands reliable and effective wireless communication for the high safety. In railway communications, different business data is transmitted through the wireless communication system for information exchange between the ground control center and the train control system. The train video monitoring system, the train passenger information system, and the wireless communication system carrying the train control information, monitoring information, and video and other multimedia information need to be supported with reliable and real-time transmission. In order to ensure the safe and effective operation of high-speed railway communication and provide a high level of service quality, wireless communication systems need to have good system performance. It is necessary to rely on the radio resource management technology which meets the characteristics of railway communication to ensure the effective transmission of all the data.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 646–654, 2019.
https://doi.org/10.1007/978-981-13-6264-4_77

At present, many researches focus on the radio resource scheduling of LTE (long-term evolution) technology under the scenario of high-speed railway. The paper [2] combines the characteristics of LTE technology on high-speed railways, analyzes the classical scheduling algorithms, and proposes a P-PF (proportional fair) algorithm that is suitable for radio resource scheduling algorithms in high-speed railway operation scenarios. During the process, the number of high-speed train users was added as the scheduling algorithm parameters to improve user priority and improve user fairness in the system. The literature [3] studied the application of the dynamic scheduling mechanism of wireless resource in OFDM (orthogonal frequency division multiplexing) system in high-speed railway communication system. In order to enable the traffic scheduling of high-speed railway communication application, a cross-layer joint resource scheduling based on OFDM system was proposed. The paper [4] studied the issue of radio resource allocation for OFMD systems in high-speed railways, taking into account the effects of Doppler shifts and the interference power between different subcarriers in high-speed mobile environments, ensuring that the total transmit power is constant. In order to maximize the throughput of the wireless system, a mixed integer programming problem was modeled for the allocation of multiple wireless resources, and the power, subcarriers, time slots, and antennas were jointly allocated. For the special scenario of LTE technology applied to high-speed railways and the business features of VoIP (voice over Internet protocol), the authors of [5] proposed a wireless resource scheduling algorithm based on adaptive modulation and coding strategies based on the idea of packing. For the characteristics of VoIP services and FTP services, a scheduling priority algorithm based on signal-to-interference plus noise ratio is proposed.

According to the characteristics and requirements of high-speed railway communication systems, there are certain differences between high-speed railway services and cellular network services. For example, video services are GBR services in cellular networks and non-GBR services in railway networks and the current resource scheduling algorithm for high-speed railway communication has less research on the business. Since the network is used in high-speed railway communication systems, the LTE-R system provides CBTC services in the downlink, UrgentText, and PIS (passenger information system) services require high real-time performance and reliability, and the requirements for rate and buffer traffic are very high. Therefore, this paper designs a wireless resource scheduling algorithm based on LTE-R system service buffer to meet the special scenario and special service requirements of high-speed railway [6].

From the high-speed rail business priority, we can see that the following table lists each type of service demand [7]. Compared with the cellular network users, the video is non-GBR service in the rail traffic business, which does not require high real-time and bit rate. Due to the special nature of the service, in the high-speed rail network, the GBR rate of the GBR service is easier to meet in the railway network, it will cause the waste of RB resources by dividing the RB into the RB business.

This paper is organized as follows. In Sect. 2, we introduce the scenarios we used in this paper and proposed the new resource scheduling algorithm. In Sect. 3, the proposed algorithm is simulated and compared with the traditional RAD algorithm. The conclusions are given in Sect. 4.

2 System Model and Algorithm

In this paper, we consider the downlink multi-bearer scenario of a single high-speed railway subscriber under private network conditions. The specific parameters of the services and services it carries are as follows [8]:

Table 1. The services in downlink of LTE-R

QCI	ServiceType	Priority	Service	PacketlossRate
2	GBR	4	CBTC	1.0e-3
2	GBR	4	UrgentText	1.0e-3
6	Non-GBR	6	PIS	1.0e-6

The scheduling is divided into time domain scheduling and frequency domain scheduling. The time domain scheduling sorts the services in the time domain scheduling list according to a certain priority order and selects the highest priority services Nmax. If the number of schedulable services is less than Nmax, then all Select and store these services in the frequency domain scheduling list. The frequency domain scheduler will allocate resources according to the order of the services in the time domain scheduling list.

The decision factor for determining whether the service n can be sent to the frequency domain scheduling at the current TTI is [1]

$$M^{TD}[n,t] \;=\; \tau[n,t] \cdot RA^{traf}[n,t] \cdot DS^{traf}[n,t]$$

indicating the time domain priority of service n at TTI t. Put the former Nmax services into the frequency domain scheduling list.

$\tau[n,t]$ is an accumulation counter, which represents the number of times that UE n has not been scheduled in the time domain at the time of TTI t.

$DS^{traf}[n,t]$ is a function of delay. Once service n has a larger scheduling delay in the packet data queue head at TTI t, the value will increase, and the value of $M^{TD}[n,t]$ will increase accordingly, making it easier for the service to be scheduled in the time domain.

$RA^{traf}[n,t]$ is the request activation value, which has different calculation methods according to different service attributes. It describes the time when service n requests for sharing at TTI t. In the view of the time domain scheduler, the maximum time domain share is Nmux, the maximum number of schedulable services.

The time domain scheduling list is ranked according to the time domain priority calculated by the service, and the top Nmax services are selected and placed in the frequency domain scheduling list as the scheduling object of the RAD frequency domain scheduling. The number of RBs required by the service in the GBR List is determined according to the respective GBR requirements in combination with the channel conditions. The number of RBs required by service n at time t $N\left[n\right]\left[t\right]$ is

$$N\left[n\right]\left[t\right] = \left\lceil \frac{LeftBit_n^t / WCQI}{CQIModulation * RENum * RI} \right\rceil$$

$LeftBit_n^t$ is the number of bits that user n still needs to send at time t.$WCQI$ is Wideband CQI. RI is RI instructions.

$$LeftBit_n^t = \log_2^{\left(\log_2^{GBR*(CurrentTTI-ServiceBeginTTI)-PHYCorrectReceiveBits}\right)}$$

where $CurrentTTI$ is the current TTI number, $ServiceBeginTTI$ is the service initiation time, and $PHYCorrect$R$eceiveBits$ is the number of correctly transmitted MAC layer bits.

The user priority list obtained by time domain scheduling is divided into a GBR List and an N-GRB List. It is necessary to separately determine the number of RBs required by each service in the two linked lists. First, the number of RBs required by the GBR is determined, and then the remaining resources are equally distributed to the N-GBR. The determination of the number of RBs for GBR users is performed as in Fig. 1.

If $GBR \times (CurrentTTI - ServiceBeginTTI) - PHYCorrectReceiveBits \geq 0$, the service cannot meet the GBR requirement, and the number of RBs allocated to the GBR service is

$$N_{GBR-RB} = \sum_{n \in GRBList} N\left[n\right]\left[t\right]$$

If $GBR * (CurrentTTI - ServiceBeginTTI) - PHYCorrectReceiveBits \prec 0$, the service can meet the GBR requirement, then treat the service as an non-GBR service, and allocate the remaining resources according to the service buffer ratio of the current TTI. The total number of remaining RBs available for non-GBR allocation is

$$N_{Non-GBR-RB} = N_{RB} - N_{GBR-RB}$$

The number of RBs allocated for each non-GBR service is

$$N_{Non-GBR-RB} = (N_{RB} - N_{GBR-RB}) \times \frac{Currentbufferi}{\sum_N Currentbufferi}$$

where $Currentbuffer$ is the buffer of non-GBR i in the current TTI.N is the number of all the non-GBR service in the current TTI.

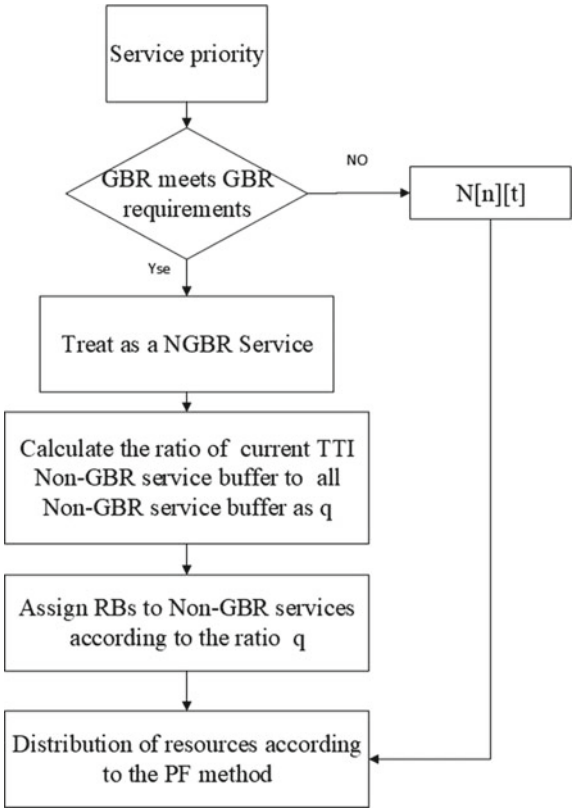


Fig. 1. Algorithm flowchart

After determining the number of RBs required by each services, use the PF algorithm's scheduling criteria, then services in the GBR List are preferentially allocated, and services in the non-GBR List are then assigned.

3 Simulation Results

The previous section designed the uplink service scheduling strategy of the LTE-R system. In order to verify the suitability of the scheduling strategy for the downlink services integrated in the LTE-R system and the system performance impact, the service carried on the LTE-R is combined with the foregoing uplink. The resource scheduling strategy is validated on the high-speed mobile simulation platform of LTE-R. The settings of the main simulation parameters are the same as the parameters of the downlink scheduling simulation [9] (Table 1).

Considering the user bear three different types of downlink services, CBTC services, UrgentText services, and PIS services. Service generation module generates 4 analog CBTC services, 1 analog PIS service, and 1 emergency text service.

The duration of the CBTC service is the entire simulation process. The time interval and duration of the business packet generation are shown in Table 2. The UrgentText service can be divided into ON/OFF phases. The ON phase sends a series of data packets, the sum of the data packets is the text size of the transmission. The OFF phase indicates the reading phase after the text is sent, the train performs corresponding processing according to the text information. For the PIS service, each frame of PIS data arrives at a specified time interval of 100 ms; the arrival time of the data packet is fixed at 5 ms in each frame. A data buffer queue needs to be established for each type of bearer service of the user to store the service data to be transmitted. The respective data queues are formed separately, and the data packets are scheduled in a first-in, first-out order.

Table 2. Service parameters

Service	ServiceType	Packet size	Send interval
CBTC	GBR	400 Bytes	32 ms
Urgenttext(on)	GBR	400 Bytes	32 ms
PIS	Non-GBR	1400 Bytes	5 ms

In each scheduling period, the amount of data actually transmitted by the scheduled the services in the scheduling period is obtained according to the allocated radio resources and the determined modulation and coding scheme. At the end of this scheduling period, the amount of service data actually transmitted by the dispatched service is subtracted from the buffer. The model of channel is simulated as WINNER-II [10,11]. The railway set four RRUs in a BBU. The distance between RRUs is 500 m, and the simulation duration is 80000TTI, in which the train needs to perform two handovers. The simulation of a single-cell downlink resource scheduling strategy also specifically analyzes the performance based on the average service throughput and the service average delay.

The result of user throughput simulation is shown in Fig. 2. The average throughput of the CBTC service, the urgenttext, and the PIS service is 368.4368 kbps, and 98.0823 kbps and 112.3442 kbps are improved to 404.016 kbps, 105.2354 kbps and 126.7348 kbps. With the increase in data packets, the packet throughput of CBTC services and PIS have been improved significantly. Compared with the traditional RAD algorithm, the rate of data is guaranteed. The average throughput of various services has been improved to a certain extent.

With the increase of data packets, the delay of CBTC service and emergency text service is in a stable trend. The average packet delay of the urgenttext dropped from 53.3725 kbps to 23.7903 ms. The average packet delay of CBTC services and PIS services maintain the level of the original RAD algorithm, CBTC changes from 443.6821 ms to 449.7625 ms, the PIS change from 7080.4987 ms to 7072.4594 ms. The simulation results of average delay for different services of users are shown in Fig. 3.

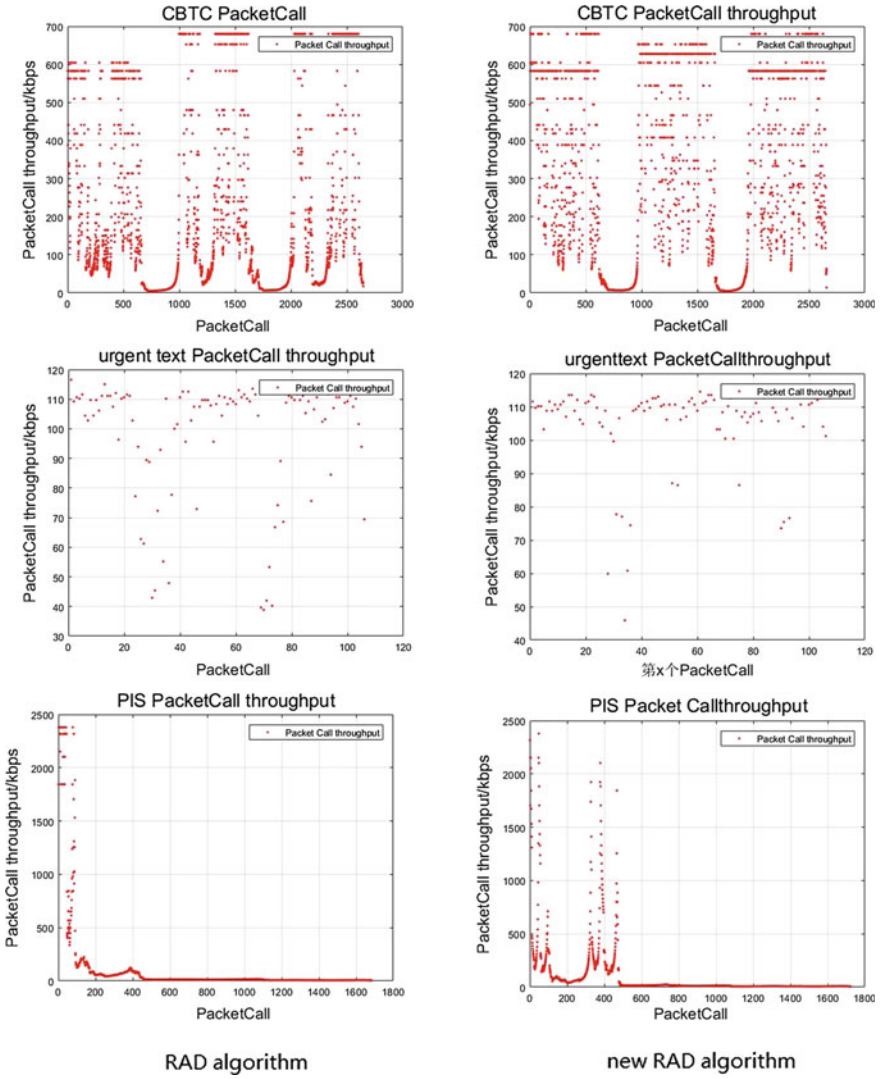


Fig. 2. Packet call throughput

Due to the nature of the service, some GBR services are treated as non-GBR services. In this way, under the traditional RAD algorithm, RBs are divided equally with non-GBR services. For GBR services, both throughput and packet delay are limited. RBs are allocated according to the buffer accumulated by the current service, which can guarantee the performance of the GBR service that is regarded as non-GBR, and it can also improve non-GBR service-related performance.

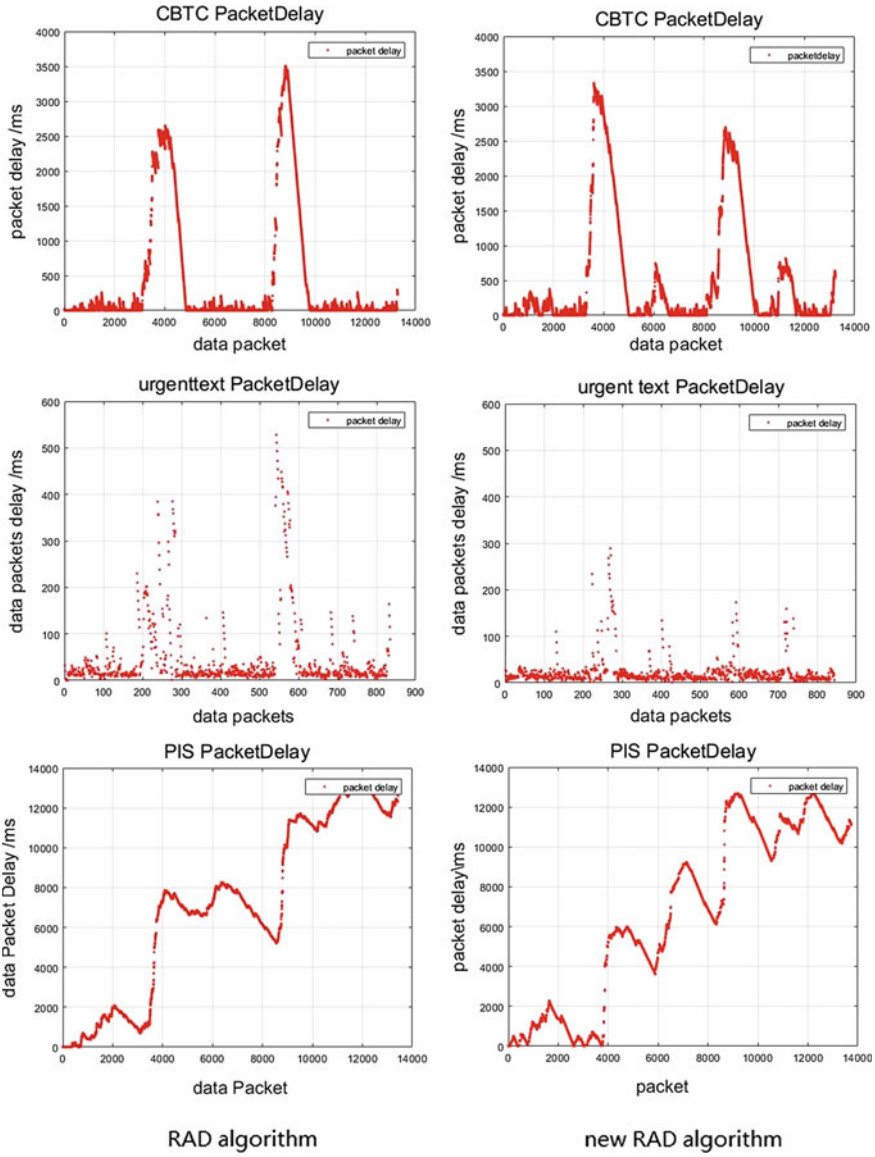


Fig. 3. Packet delay

4 Conclusion

In this paper, we propose a resource scheduling algorithm that allocates RBs based on the ratio of current TTI traffic buffers. Our algorithm focuses on non-GBR services and the GBR services which are regarded as non-GBR services. The traditional RAD algorithm distributes the RBs of the non-GBR service

evenly. Under the special scenario of high-speed trains, the PIS, a non-GBR service, is divided into RBs that are relatively low relative to the generated rate. However, to the GBR service, the throughput and packet delay also have a large room for improvement. In the LTE-R dynamic simulation platform, we simulate the average packet throughput and the average packet delay of CBTC service, the urgent text and the PIS service. The simulation results show that some indexes improved more clearly. The results show that the resource algorithms based on the improved RAD algorithm proposed in this paper basically meet the communication requirements of LTE-R wireless communication system.

There are several extensions to this work and we hope to return in the future. For example, considering the special scene of high-speed railway, we will set the threshold of packet delay to ensure that all data packets are sent within a certain period of time to meet the high-speed rail reliability requirements. According to the new RAD algorithm proposed in this paper, the allocation of RBs is performed in a certain proportion. This allocation may reduce certain performance compared with the algorithm of this paper, but it can reduce the signaling interaction and reduce the complexity of the algorithm.

Acknowledgement. This work described in this paper was supported by the National Science and Technology Major Project of the Ministry of Science and Technology of China (Grant No. 2016ZX03001009-003).

References

1. Monghal, G., Laselva, D., Michaelsen, P.H., et al.: Dynamic packet scheduling for traffic mixes of best effort and VoIP users in E-UTRAN downlink. In: Vehicular Technology Conference. IEEE Xplore, pp. 1–5 (2010)
2. Yan, X., Lu, L.: Research on LTE radio resource scheduling algorithm based on high speed railway. J. China Railw. Soc. **38**(09), 81–85 (2016). in Chinese
3. Anming, J.: Application of OFDM System Resource Scheduling and Allocation Strategy in Railway Communication System. Beijing Jiaotong University (2009)
4. Yisheng, Z.: Research on Resource Management Algorithms of High-Speed Railway Broadband Communication Based on TD-LTE. Beijing University of Posts and Telecommunications (2013)
5. Xuan, Y.: Resource Scheduling Algorithm for Co-existence of VoIP Service and Multi-services in High-speed Rail Wireless Communication. Southwest Jiaotong University (2015)
6. 3GPP TS 36.203. Tech. Specif. Group Services and System Aspects-Policy and charging control architecture (Release 9)
7. Urban Rail Transport Association Technical Equipment Committee. CB/T.V0.3-2015.LTE - System Requirements Specification **10**, 9–12 (2015). in Chinese
8. 3GPP TS 23.203 V9.12.0 (2012-06) Policy and charging control architecture
9. Sesia, S., et al.: Long-Term Evolution of LTE-UMTS Theory and Practice. Beijing, People's Posts and Telecommunications Press. in Chinese
10. 3GPP TS 36.211 V8.9.0 (2009–12). Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation
11. 3GPP TS 36.212 V8.8.0 (2009-12). Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and channel coding

A Handover Optimization Algorithm for LTE-R System Handover Parameter Prediction and Dynamic Adjustment

Shuai Zhang(✉), Gang Chuai, and Weidong Gao

Ministry of Education, Key Laboratory of Universal Wireless Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China
{jason_zhang,chuai,gaoweidong}@bupt.edu.cn

Abstract. Handover is one of the key technologies to guarantee the quality of high-speed rail wireless communication. The existing LTE-R handover algorithm has the following deficiencies: 1. The same channel model is adopted for different speed trains; 2. the fixed trigger time (TTT) and hysteresis are used. Thresholds, or segmented adjustments to only the hysteresis thresholds, cannot meet the changing high-speed rail speed requirements for handover performance. In this paper, the following improvements are provided: 1. Different channel models are used for different train speeds in the handover process to more accurately reflect the status of wireless channels at the time of handover. 2. Inverse function is used to establish the speed and trigger time based on the least squares method. The optimal handover parameter configuration at different speeds can be predicted by establishing a functional relationship between (TTT) and the hysteresis threshold. The proposed optimization algorithm based on dynamic handover parameters provides an important reference for the setting of handover parameters for high speed railways at different speeds. The simulation results show that in the LTE-R system, the proposed algorithm can better predict the handover parameter configuration at different speeds, and the handover success rate and the wireless link failure rate index are improved.

1 Introduction

The establishment of high-speed rail communication technology must not only meet the needs of train passengers for mobile communications but also provide security for the high-speed rail trains. As an important technology in high-speed rail communications, handover technology has become a research hotspot for high-speed rail communications. The literature [1] introduces and simulates the handover algorithm and handover flow based on A3 events in traditional LTE, but it cannot meet the high-speed rail scenario. Demand; Literature [2] evaluated handover performance by selecting different TTT values for different UE speeds and neighboring cell types in an SON-based LTE system. Literature [3] proposed weighted performance based on switching parameter optimization

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 655–669, 2019.
https://doi.org/10.1007/978-981-13-6264-4_78

methods. The work in [4] evaluates the applicability of RSRQ quality measurement based on different frequency handover. Documents [5,6] proposed some new handover algorithms based on geographic location information and UE mobility history. The work in [7,14] introduces the handover algorithm in self-organizing networks. In [9,11], a handover algorithm based on distance and RSSI measurement is described. However, due to ignoring the influence of speed, the handover performance is still insufficient. The literature [8,10,15,16,18] mainly studies and discusses the optimization schemes of handover algorithms in high-speed rail communication at different angles. And the literature [12,13,17,20] proposes new research ideas in the configuration of handover parameters. All of the above studies have not considered the situation that the high-speed rail train changes with the change of speed of the wireless channel. Fixed selection of handover parameters also makes it impossible for the handover parameters to provide dynamic adjustment schemes for different train speeds. These handover algorithms are in the rapidly deteriorating chain. The road quality and wireless channel environment are still deficient in high-speed railway scenarios that become very unstable due to increased speed.

Therefore, the innovations of this paper are as follows: 1. For different speed conditions in high-speed rail scenes, different channel models are used to simulate the true high-speed rail wireless environment, so that the performance of the handover algorithm is most verified. 2. According to the relationship between the optimal handover parameters and train speed, an inverse function is proposed as a dynamic prediction function to predict the configuration of handover parameters at different speeds. 3. An optimized handover algorithm based on dynamic handover parameters is proposed to provide dynamic handover parameter configuration for high-speed rail handover at different speeds to provide reliable handover for high-speed rail wireless communication systems.

The arrangement of this paper is as follows. In the second section, the basic concepts of LTE-R handover and the flow of handover are mainly introduced. In the third section, a high-speed railway distributed base station deployment scheme for multiple RRU common cells is introduced. In the fourth section, based on the inverse function, an improved handover algorithm for high-speed railway handover parameters prediction and dynamic handover parameters is proposed. The fifth section carries on the simulation verification to the proposed algorithm. Finally, the conclusions and future work are discussed in Section V.

2 LTE-R Handover and Challenges

2.1 Traditional Handover Algorithm Based on A3 Events

When the conditions for entering the A3 event are met, the UE measures the signal quality of the current serving cell and the target cell with a layer-three filtering period T_u, if the measured value satisfies the A3 event during the switching delay time TTT, the measured value is reported; otherwise, the measured value is returned to the measurement phase that does not enter the A3 event. If the

measured value satisfies the A3 event, the measured value is reported; otherwise, it returns to the event that does not enter the A3 event.469681 469684

The switching conditions based on A3 time are as follows:

$$RSRP_t - RSRP_s \geq Hysteresis + OFF$$

$$T \geq TTT$$

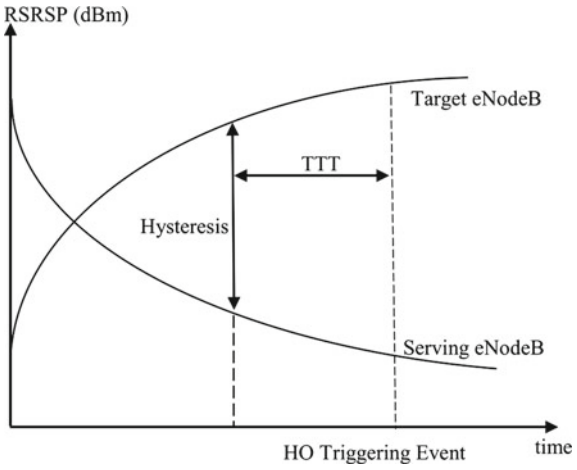


Fig. 1. A3 event trigger-based handover algorithm

2.2 High-Speed Rail Switching Challenges

High-speed rail communication has features such as high body penetration loss, serious Doppler effect, complex wireless channel, and frequent switching. Due to the high-speed rail scene, trains run fast (the fastest speed of Chinese Fuxing train reaches 350 km/h). However, in order to ensure good coverage of the base station, the base station spacing is limited. Therefore, when the mobile terminal moves quickly, the handover needs to be completed in a very short time, so as to ensure that the high-speed rail operation stage can achieve a highly efficient and reliable communication experience. Therefore, the reasonable setting of the handover parameters plays a crucial role in high-speed rail communications (Fig. 1).

3 HIGH-Speed Railway Distributed Base Station Deployment Plan

3.1 Concept

The main idea of the distributed base station is to spatially separate the baseband processing unit BBU (building baseband unit) of the traditional macro

base station from the radio frequency processing unit RRU (remote radio unit). When the network is deployed, the BBU and the core network and the wireless network control equipment are centralized in the equipment room, and multiple optical fibers are used. The RRU is connected to the same BBU and maximizes resource sharing. This improves the quality of wireless communications and facilitates the construction, operation, and maintenance of the base station (Tables 1, 2 and 3).

3.2 High-Speed Railway Distributed Base Station Deployment Plan

The deployment of high-speed railway distributed base stations mainly uses the technology of multiple RRU common BBUs to achieve network coverage along high-speed railways. Its main concept is to continuously deploy RRUs belonging to the same BBU along the line and set the RRU devices at different locations to the same frequency through automatic frequency correction technology, that is, multiple RRUs theoretically belong to the same cell. Through the multi-RRU common cell technology, the coverage area is expanded, and a reasonable network architecture solution is provided under the special scene of high-speed rail (Fig. 2).

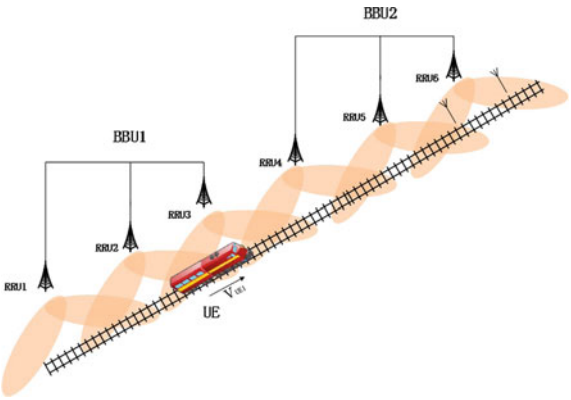


Fig. 2. High-speed rail distributed base station deployment

For the high-speed railway distributed deployment scheme, its main advantages are as follows:

1. Multiple RRU common cells can increase the coverage of a single base station, thereby reducing network construction costs.
2. Since multiple RRUs are managed by the same BBU, the handover is performed only between the two BBUs, and the signal continuity of the HSR can be ensured while the number of handovers is less.

4 Handover Parameter Optimization and Dynamic Adjustment Handover Optimization Algorithm

4.1 Simulation is Based on Different Channel Models

In [19], the mountainous scene using the WINNER II channel model predicts the handover parameters from 0 km/h to 350 km/h, but this approach is not suitable. The reasons are as follows: 1. At medium and low speeds, the channel modeling method in the WINNER II channel model does not match the channel environment in high-speed trains with low speeds. 2. At low and medium speeds, the communication system will use large hysteresis thresholds and TTT to avoid ping-pong handovers, but ping-pong handovers are not prone to occur at high speeds.

Table 1. Changes of optimal hysteresis threshold at 50 km/h

Configuration (dB ms)	Handovers initiated	Successful handovers	RLF (radio link failure)	Success rate (%)
1,600	241	188	53	78.00
2,600	198	157	41	79.29
4,600	233	180	53	77.25
6,600	187	136	51	72.73

Table 2. Changes of optimal hysteresis threshold at 120 km/h

Configuration (dB ms)	Handovers initiated	Successful handovers	RLF (radio link failure)	Success rate (%)
1,600	302	236	66	78.15
2,600	316	225	91	71.20
4,600	403	295	108	73.20
6,600	387	274	113	70.80

Therefore, in order to better guarantee the effectiveness of the handover algorithm and better reflect the real high-speed rail wireless communication environment, we use different channel models for different speed high-speed trains, i.e., medium-speed and low-speed scenarios ($V \leq 120$ km/h) adopt the ITU-VA channel model, and high-speed scenario ($V \geq 120$ km/h) adopts the WINNER II channel model, so that the simulation scenario can best meet the actual situation.

4.2 Use an Inverse Function that More Closely Matches the Change in Handover Parameters

The goal of this paper is to predict the handover parameters at different speeds by using existing data and to increase the handover success rate by dynamically adjusting the parameters. In [19], the relationship between train speed and handover parameters (hysteresis threshold and TTT) was fitted and proved by using elliptic function. However, because it uses the WINNER II channel at different speeds and does not conform to the actual scenario, we fit the speed and switching parameters based on the channel models at different speeds. The fitting results are as follows:

(1) Changes Of Optimal TTT at Different Speeds
According to the simulation results, it can be seen that as the train speed increases, the hysteresis threshold and TTT gradually decrease. In addition, the rate of change of the switching parameters also gradually decreases, which is consistent with the inverse characteristics of the function, so we choose the inverse function as the switching parameter. Dynamic prediction function (Tables 4, 5, 6, 7, 8, 9, 10 and 11):

$$y = \frac{1}{ax + b}$$

Since the handover parameters (hysteresis threshold and TTT) are consistent with the inverse function as the train speed changes, by substituting the hysteresis threshold into the function y, the functional relationship between speed and hysteresis threshold is

$$H = \frac{1}{av + b}$$

Table 3. Changes of optimal hysteresis threshold at 200 km/h

Configuration (dB ms)	Handovers initiated	Successful handovers	RLF	Success rate (%)
0,150	189	142	47	75.13
1,150	232	166	66	71.55
3,150	175	122	53	69.71
6,150	136	97	39	64.70

Table 4. Changes of optimal hysteresis threshold at 350 km/h

Configuration (dB ms)	Handovers initiated	Successful handovers	RLF	Success rate (%)
1,150	212	107	105	50.47
3,150	164	29	135	17.68
6,150	109	7	103	6.42

Table 5. Changes of optimal TTT at 50 km/h

Configuration (dB ms)	Handovers initiated	Successful handovers	RLF	Success rate (%)
4,150	175	125	50	71.42
4,200	163	123	40	75.46
4,300	204	135	69	66.18

Using the least squares method, substituting the optimal hysteresis threshold into the formula gives

$$H1 = \frac{1}{av1 + b}, H2 = \frac{1}{av2 + b}, H3 = \frac{1}{av3 + b}, H4 = \frac{1}{av4 + b}, H5 = \frac{1}{av5 + b}$$

where $H1 = 6$, $v1 = 0$, $H2 = 2$, $v2 = 50$, $H3 = 1$, $v3 = 120$, $H4 = 0$, $v1 = 200$, $H5 = 0$, and $v5 = 350$. The algorithmic idea of the least squares estimation method is to minimize the variance of the two sides of the equal sign, that is, to find the minimum value of this function (Fig. 3):

$$S(a,b) = (6 - \frac{1}{a \times 0 + b})^2 + (2 - \frac{1}{a \times 50 + b})^2$$
$$+(1 - \frac{1}{a \times 120 + b})^2 + (0 - \frac{1}{a \times 200 + b})^2$$
$$+(0 - \frac{1}{a \times 350 + b})^2$$

Table 6. Changes of optimal TTT at 120 km/h

Configuration (dB ms)	Handovers initiated	Successful handovers	RLF	Success rate (%)
4,150	377	284	93	75.33
4,300	463	343	120	74.08
4,600	403	295	108	73.20

Table 7. Changes of optimal TTT at 200 km/h

Configuration (dB ms)	Handovers initiated	Successful handovers	RLF	Success rate (%)
0,1	185	141	44	76.21
0,85	257	197	60	76.65
0,150	189	142	47	75.13

Table 8. Changes of optimal TTT at 350 km/h

Configuration (dB ms)	Handovers initiated	Successful handovers	RLF	Success rate (%)
0,1	247	204	43	82.59
0,150	237	150	87	63.29
0,300	203	91	112	44.83

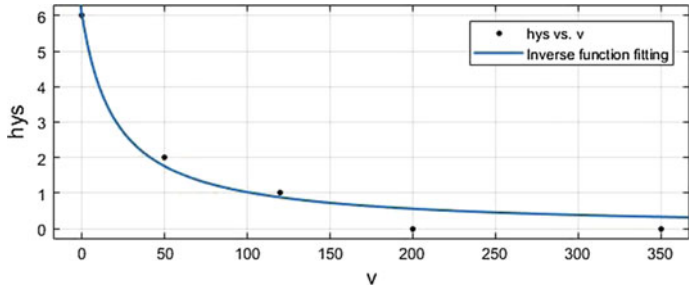


Fig. 3. Relationship between speed and hysteresis threshold

Find the partial derivative of S and let it $= 0$, that is

$$\frac{\partial S}{\partial a} = 0, \frac{\partial S}{\partial b} = 0$$

One can get: $a = 0.008065$, $b = 0.1662$.

Substituting the fitting result into the inverse function formula gives

$$H = \frac{1}{0.008065 \cdot v + 0.1662}$$

The same reason can be obtained the functional relationship between speed and TTT is

$$TTT = \frac{1}{av + b}$$

According to the least squares estimation can be obtained: $a = 5.303e - 05$, $b = 0.002107$.

Substituting the fitting result into the inverse function formula gives

$$TTT = \frac{1}{5.303\text{e} - 05 \cdot v + 0.002107}$$

4.3 Inverse Function-Based Dynamic Function Adjustment Algorithm

Step1 Selection of medium and low-speed ITU-VA and high-speed WINNER II rural macro cell channel models.

Step2 The relationship between different hysteresis thresholds, TTT, and handover success rate based on the simulation of 50 km/h and 120 km/h at medium and low-speed ITU-VA channel model (Fig. 4).

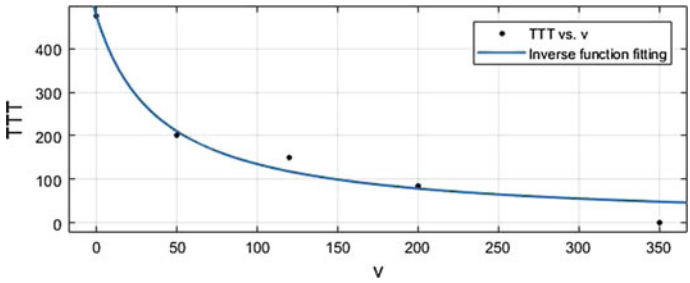


Fig. 4. Relationship between speed and TTT

Table 9. Shadow fade in VA scene

Scenes	SF (dB)	Range and antenna default height
NLOS	$\sigma = 10$	0m<d<50m, $h_{bs} = 15$ m

Table 10. Basic module path loss model summary

Path loss	SF standard deviation	Range and antenna height
		$30m < d < d_{BP}$
A=21.5 B=44.2 C=20	$\sigma = 4$	$d_{BP} < d < 10$ km
PL^*	$\sigma = 6$	$h_{BS} = 32$ m
		$h_{MS} = 1.5$ m

Step3 The relationship between different hysteresis thresholds, TTT, and handover success rate based on the high-speed WINNER II channel model simulation at 200 km/h and 350 km/h speed.

Step4 Fitting of handover parameters and train speed by least square method.

Step5 Get the functional relationship between the train speed and the handover parameters (hysteresis threshold and TTT).

Step6 Prediction of handover parameters at other speeds based on inverse function relations.

Step7 Use different handover parameters for different speed trains to achieve the goal of increasing the success rate of handover.

4.4 Channel Model

ITU-VA Channel Model. A: Path loss

$$L = 40\ (1 - 4 \times 10^{-3}\,\Delta h_b)\ \log_{10} R - 18\ \log_{10} \Delta h_b + 21\ \log_{10} f + 80\,\text{dB}$$

where R is base station to mobile station distance (km), $f = 2000\,\text{MHz}$, h_b is the antenna height of the base station (typically fixed at 15 m, with a range of 0 to 50 m).

B: Shadow fade In the VA scenario:

Table 1 lists the shadow fading in the NLOS scenario under this propagation scenario.

WINNER Channel Model. A: Path loss In D2a scenario: The path loss of the WINNER II channel model is

$$PL^* = 40.0\log_{10}(d) + 10.5 - 18.5\log_{10}(h_{MS}) + 1.5\log_{10}(\frac{f_c}{5.0})$$

where d is the distance from the transmitter to the receiver,f_c is the system frequency.

B: Shadow fade

5 Simulation

5.1 Simulation Parameter Configuration

In this work, the simulation configuration of the improved algorithm is as follows:

5.2 Simulation Results

Changes of Optimal Hysteresis Threshold at Different Speeds In this part, we predict the hysteresis threshold and TTT at 80 km/h and 250 km/h based on the functional relationship between speed and handover parameters to prove the correctness of the dynamic prediction function. Substituting speed into hysteresis threshold dynamic prediction function, you can get: when $V = 80$ km/h, the optimal hysteresis threshold = 4dB (TTT = 150 ms). The simulation results are as follows (Figs. 5, 6, 7, 8 and 9):

From the simulation results, we can see that, based on the handover parameter configuration predicted by the optimization algorithm (hysteresis threshold = 4 dB), the handover success rate is higher than that of other configurations, indicating that the predicted handover parameters can ensure better handover.

Substituting speed into TTT dynamic prediction function, you can get: when $V = 80$ km/h, the optimal TTT = 60 ms (hysteresis threshold = 1 dB). The simulation results are as follows:

Table 11. Simulation parameter configuration

Simulation parameter	Parameter configuration
Base station power P_{eNB}	49 dbm
Transmit gain G_b	17 dbi
Accept gain G_m	0 dbi
Base station height H	35 m
Train height h	3.6 m
Base station frequency f	450 M/1.8 G
Band B	5 M
RRU count of the same BBU	4
RRU spacing D	500 m
Fastest train speed V_{max}	350 km/h
Channel model M	WINNER /ITU-VA
Number of simulated Drops	1000
Number of simulated TTIs	10000
RSRP measurement period	150 ms
RSRP reporting period	300 ms

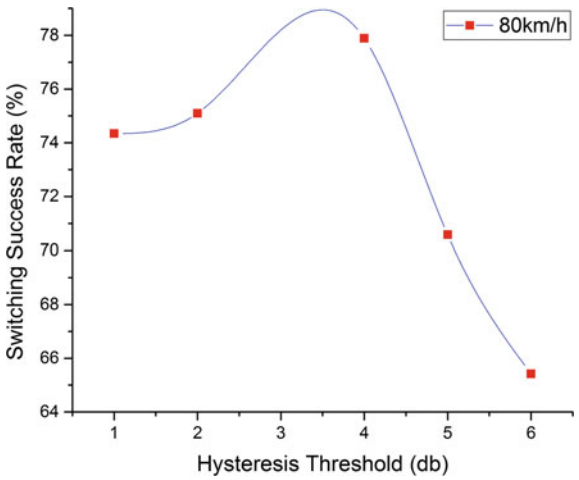


Fig. 5. Hysteresis threshold and switching success rate in speed = 80 km/h

From the simulation results, the handover success rate corresponding to the predicted handover parameter (TTT = 600 ms) is 74.04%. Although it is slightly lower than TTT = 300 ms, it is still the second best handover success rate at this speed.

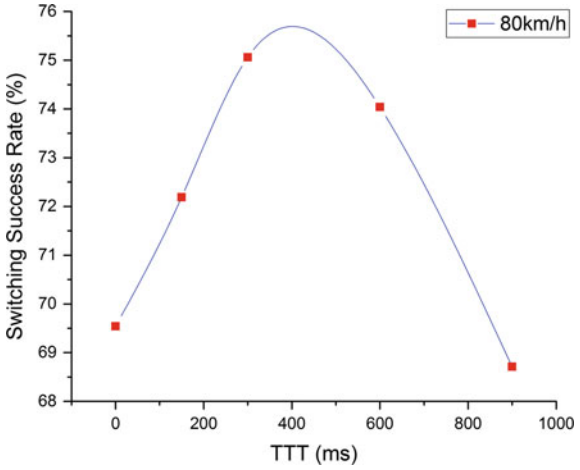


Fig. 6. Relationship between TTT and handover success rate at speed = 80 km/h

When $V = 250$ km/h, the optimal hysteresis threshold = 0 dB (TTT = 150 ms):

From the simulation results, we can see that, based on the handover parameter configuration predicted by the optimization algorithm (hysteresis threshold = 0 dB), the handover success rate is higher than that of other configurations, indicating that the predicted handover parameters can ensure better handover.

When $V = 250$ km/h, the optimal TTT = 75 ms (hysteresis threshold = 0 dB):

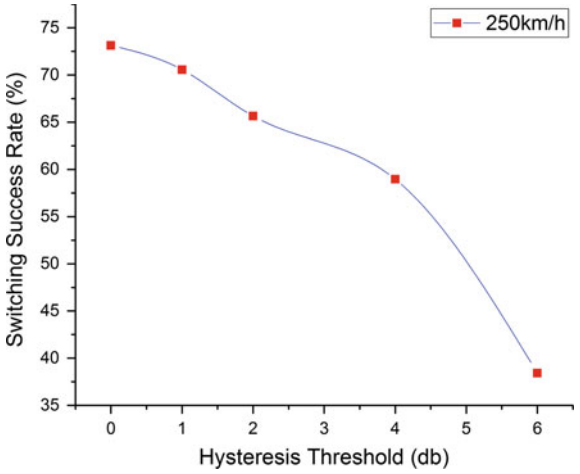


Fig. 7. Hysteresis threshold and switching success rate in speed = 250 km/h

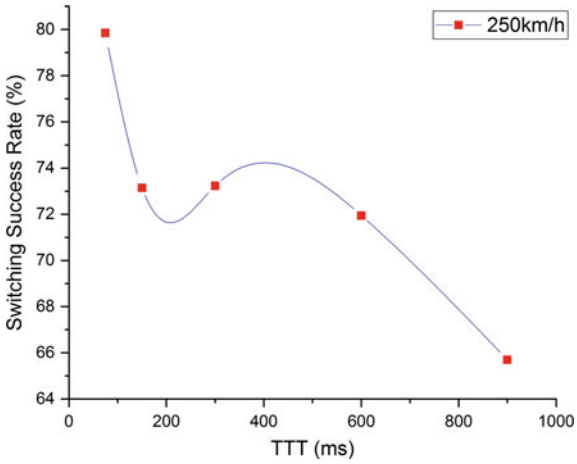


Fig. 8. Relationship between TTT and handover success rate at speed = 250 km/h

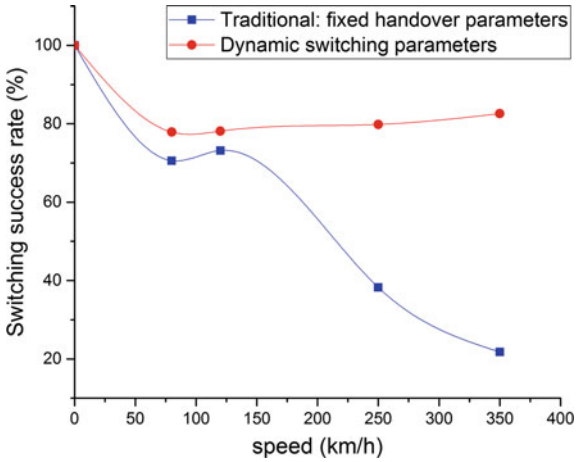


Fig. 9. Comparison of fixed handoff parameters and dynamic parameter algorithms

It can be seen from the simulation results that the handover success rate based on the prediction after the optimization algorithm is configured (TTT = 75 ms), the handover success rate is higher than that of other configurations, indicating that the predicted handover parameters can guarantee the handover success.

As shown in the figure, the traditional handover algorithm has a fixed hysteresis threshold = 4dB, TTT = 600 ms, and the handover success rate is as shown in the figure. As the speed increases, the handover success rate of the fixed handover parameter decreases sharply. When the speed is equal to 350 km/h.

The handover success rate is less than 30%, and the improved handover algorithm based on dynamic handover parameters can better improve the success rate of high-speed rail handover, making it always above 77%.

6 Conclusion and Future Work

This paper mainly discusses the problem of handover parameter configuration based on high-speed railway scenarios. Compared with the traditional handover algorithm based on fixed handover parameters of A3 events, this paper establishes the inverse function dynamic prediction relationship between speed and hysteresis threshold and TTT. In addition, the dynamic adjustment algorithm based on different handover parameters given at different speeds can provide the best handover parameters for trains with different speeds in the high-speed rail scene. From the simulation results, the proposed algorithm can predict the handover parameters to some extent and improve the success rate of handover.

In the future work, one can consider combining the optimal hysteresis threshold and TTT, on the one hand, to find the best match of dynamic handover parameters at different speeds and, on the other hand, consider using different functional relationships to perform the best handover parameters. Find a more accurate fit function through multiple fittings to provide a better handover scheme for high-speed rail handover.

Acknowledgments. This work described in this paper was supported by the National Science and Technology Major Project of the Ministry of Science and Technology of China (Grant No. 2016ZX03001009-003).

References

1. Rick, M.R.M., Badran, E.F.: Study of LTE-R X2 handover based on A3 event algorithm using MATLAB, pp. 1155–1159 (2015)
2. Lee, Y., Shin, B., Lim, J., Hong, D.: Effects of time-to-trigger parameter on handover performance in SON-based LTE systems. In: Asia-Pacific Conference on Communications, 2010. APCC 2010, pp. 492–496. IEEE (2010)
3. Jansen, T., Balan, I., Stefanski, S., Moerman, I., Kurner, T.: Weighted performance based handover parameter optimization in LTE. In: Vehicular Technology Conference: VTC- 2011 Spring. IEEE 73th, pp. 1–5. IEEE (2011)
4. Kurjenniemi, J., Henttonen, T., Kaikkonen, J.: Suitability of RSRQ measurement for quality based inter-frequency handover. In: IEEE International Symposium on LTE, in Wireless Communication Systems: ISWCS 2008, pp. 703–707. IEEE (2008)
5. Ge, H., Wen, X., Zheng, W., Lu, Z., Wang, B.: A history-based handover prediction for LTE systems. In: IEEE 1th International Symposium on Computer Network and Multimedia Technology: CNMT 2009, pp. 1–4. IEEE (2009)
6. Markopoulos, A., Pissaris, P., Kyriazakos, S., Sykas, E.D.: Efficient location-based hard handoff algorithms for cellular systems. In: 3rd International IFIP Networking Conference, vol. 3042, pp. 476–489 (2004)

7. Amirijoo, M., et al.: Use cases, requirements and assessment criteria for future self-organising radio access networks. In: Proceedings of 3rd International Workshop on Self-Organizing Systems, IWSOS 08, Vienna, Austria, December 10–12, 2008
8. Zhang, N., Holtzman, J.M.: Analysis of handoff algorithms using both absolute and relative measurements. IEEE Trans. Veh. Technol. **45**(1), 174–179 (1996)
9. Itoh, K.-I., Watanabe, S., Shih, J.-S., Sato, T.: Performance of handoff algorithm based on distance and RSSI measurements. IEEE Trans. Veh. Technol. **51**(6), 1460–1468 (2002)
10. Li, Y., Yan, Y.S.: Handoff decision algorithm with dynamic sampling distance in GSM-R network. In: Proceedings IEEE International Conference Railway Engineering Challenges Railway Transportation Information Age, Hong Kong, pp. 1–4 (2008)
11. Xu, P., Fang, X., He, R., Xiang, Z.: An efficient handoff algorithm based on received signal strength and wireless transmission loss in hierarchical cell networks. Telecommun. Syst. **52**(1), 317–325 (2013)
12. Lee, D.-W., Gil, G.-T., Kim, D.-H.: A cost-based adaptive handover hysteresis scheme to minimize the handover failure rate in 3GPP LTE system. EURASIP J. Wireless Commun. Netw. (2010), Art. ID 750173
13. Lee, W., Cho, D.-H.: Enhanced group handover scheme in multiaccess networks. IEEE Trans. Veh. Technol. **60**(5), 2389–2395 (2011)
14. Sinclair, N., Harle, D., Glover, I.A., Irvine, J., Atkinson, R.C.: An advanced SOM algorithm applied to handover management within LTE. IEEE Trans. Veh. Technol. **62**(5), 1883–1894 (2013)
15. Park, J., Lim, Y.: A handover prediction model and its application to link layer triggers for fast handover. Wireless Pers. Commun. **52**(3), 501–516 (2010)
16. Zhang, Y., Wu, M., Ge, S., et al.: Optimization of time-to-trigger parameter on handover performance. In: LTE High-Speed Railway Networks International Symposium on Wireless Personal Multimedia Communications, pp. 251–255 (2012)
17. Hao, C.Y., Liu, H.I., Zhan, J.A.: Velocity-adaptive handover scheme for mobile Wi MAX. J. IEEE J. Commun. Netw. Syst. Sci. (9), 874–878 (2009)
18. Lin, C., Sandrasegaran, K. et al.: Optimization of handover algorithms in 3GPP long term evolution system C modeling. In: Simulation and Applied Optimization (ICMSAO) 2011 4th International Conference, pp. 1–4 (2011)
19. Li, D.W., Cheng, Y.G., et al.: LTE-R handover optimization algorithm based on velocity dynamic function. J. Railw. Stand. Des. (11), 143–147 (2016)
20. Luo, W.T., Fang, X., MCheng, M., et al.: An optimized handover trigger scheme in LTE systems for high speed railway C. In: 15th International Workshop on Signal Design and its Applications in Communications, pp. 193–196 (2011)

A High-Performance Synchronous Energy-Saving Algorithm for Wireless Sensor Networks

HanZhi Hu, ZhiLu Wu(✉), ZhuXi Li, and XiaoJun Zhang

School of Electronics and Information Engineering,
Harbin Institute of Technology, Harbin 150006, China
wuzhilu@hit.edu.cn,
{821771303,51280684,2378515255}@qq.com

Abstract. Time synchronization technology is an important prerequisite for effective application of wireless sensor networks, and provides basic services for multiple application services. However, due to the complexity of the network environment, it is difficult to meet the different applications of the time synchronization mechanism in practical applications. In this paper, based on the characteristics of the wireless sensor network and the application of the related time synchronization technology, a new high-performance synchronous energy-saving algorithm is proposed on the basis of the existing technical achievements. This algorithm synthetically considers the residual energy level and node degree of the child nodes, and optimizes the selection process of the reference nodes, thus reducing and balancing the energy consumption between the nodes and effectively prolonging the network life. Simulation results show the effectiveness of the algorithm.

Keywords: Wireless sensor networks · Time synchronization · Energy consumption

1 Introduction

Time synchronization of wireless sensor networks (WSN) refers to the process that each independent node exchanges local clock information with other nodes to achieve and maintain a coordinated global time [1]. Whether it is data fusion of sensor, the positioning of sensor node itself or collaborative work of sensor node, time synchronization between nodes is required. Due to the application of wireless sensor network node and the node's simplicity, wireless sensor network not only requires high synchronization accuracy, but also has a low network overhead and well extensibility.

In recent years, with the untiring research of scholars all over the world, a variety of WSN time synchronization algorithms are proposed. S. ping, in [2] proposes a synchronization model based on sender. Typical algorithms include Delay Measurement Time Synchronization (DMTS) algorithm based on time delay measurement and Flooding Time Synchronization Protocol (FTSP) algorithm based on flooding. The synchronization between nodes can be realized by sending a broadcast message. In [3], H. Dai et al propose a hybrid algorithm, such as Hierarchical Reactive Time

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 670–677, 2019.
https://doi.org/10.1007/978-981-13-6264-4_79

Synchronization Protocol (HRTS), which uses multiple wireless channels to improve synchronization accuracy, minimize communication load and reduce energy consumption. When the i layer broadcasts synchronous beacons to the $i + 1$ level node in the control channel, it randomly specifies a $i + 1$ layers' node to switch to the time channel recovery synchronization beacon. The node selection of the reply synchronous beacon does not consider the energy of the node, which may result in the node with less residual energy acting as the node of the response synchronization beacon and die prematurely, resulting in isolated nodes, thus affecting the network life. A lot of nodes in WSN are cheap and their clocks may drift away from each other in time [4, 5]. The drift and offset between clocks need to be compensated periodically in order to keep a precise common time over the nodes and it is achieved by using time synchronization protocols.

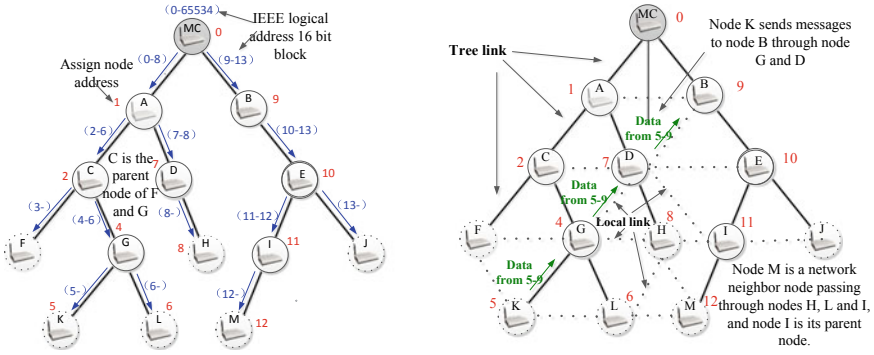
In this paper, using the idea of HRTS and Timing-sync Protocol for Sensor Networks (TPSN) time synchronization algorithm, we propose a new energy adaptive high-performance synchronization energy-saving algorithm for wireless mesh sensor networks. The algorithm selects a reference sub-node by synthetically considering the number and degree of the child nodes, and synchronizes with the parent node to achieve the purpose of reducing the information overhead and equalizing energy, so as to prolong the lifetime of the whole network.

The remainder of the paper is organized as follows. Section 2 provides the wireless network sensor network with high-performance synchronization energy-saving algorithm. Section 3 describes the analysis index of high-performance synchronization energy-saving algorithm. Section 4 gives the simulation result and the corresponding analyses. Finally, the conclusions are drawn in Sect. 5.

2 High-Performance Synchronous Energy-Saving Algorithm

In this section, the high-performance synchronization energy-saving algorithm is introduced in detail. The synchronization process is divided into two stages: The first phase is expressed as network discovery, where nodes share information about the networking of their activated operations, establish links between nodes, improving the robustness and flexibility of data transmission. The second stage indicates that nodes start to synchronize step by step when a normal operation occurs.

At the discovery stage, each node needs to know which node is its neighbor node and how far it is from the root node. A wireless node generates its own routing table through a dual program: (I) a typical tree-based addressing scheme can reduce the size of the address field in the message header. (II) a link between neighbors in the neighborhood to generate the final grid tree topology. After the root node receives the hierarchical discovery of the child node broadcast, the child node is added to its child node table. The root node is responsible for the continuous 16 bits' logical block address assigned to each node, as shown in Fig. 1a. Figure 1b shows that the route robustness process is to reduce the energy loss and network congestion caused by flooding broadcast through additional local links. This stage is mainly to establish the hierarchical structure of the whole network.



(a) allocation address (b) new generation of local links

Fig. 1. The first stage of the synchronization process: the discovery stage

At the synchronization stage, the main task of this stage is to synchronize the nodes in the network layer by layer from the root node to achieve the whole network synchronization, and to ensure the balance of the node energy consumption through the node selection algorithm, thus prolonging the network life.

In the synchronization process, Node A sends a synchronization request message to Node B at the time T_1 and Node B receives the message at the time T_2. Then at the time T_3, Node B sends back a synchronous reply message to Node A. Node A receives the synchronous reply message at the time T_4 [6–8].

$$\begin{cases} d = \frac{(T_2 - T_1) - (T_4 - T_3)}{2} \\ \Delta = \frac{(T_2 - T_1) + (T_4 - T_3)}{2} \end{cases} \tag{1}$$

In (1) some parameters must be determined, d represents clock drift between two nodes; Δ represents message round-trip delay.

As shown in Fig. 2, in the synchronization process of the node, Node A and Node B exchange a pair of unicast messages, and calculate the four different time values of Node A and Node B. Each of these times corresponds to a specific time, T_1 and T_4 represent the time of local clock measurement for node A, T_2 and T_3 represent the time of local clock measurement for node B.

The root node broadcasts the synchronous reference packet, and all the nodes that receive the synchronous reference packet record the receiving time of the message with their own local clock, but only the reference sub-node specified in the synchronization packet returns the response message.

Mesh coordinator is the head device of tree topology. It is responsible for sending messages and starting the synchronization process. When second messages arrive at the sub-nodes, they use the different clock times between the coordinators and the previously stored timestamps, using the low delay messages between the two consecutive synchronizations. The synchronization of the first region ends when the synchronous

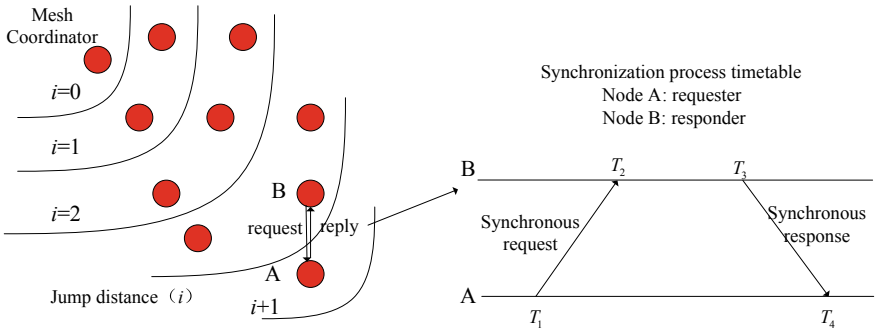


Fig. 2. Schematic diagram of information transmission

response message reaches the longest node in the first region, and the second region begins the same process.

The mesh coordinator is responsible for the duration of the entire synchronization process, as shown in Fig. 3. The energy-saving mechanism defines WO (Wake Order), AO (Active Order) and SI represents the synchronization interval. During the activity period, the energy-saving mechanism uses the end-to-end delay to propagate the message, which is minimized during the inactive period, and all nodes attempt to access the physical layer transport information.

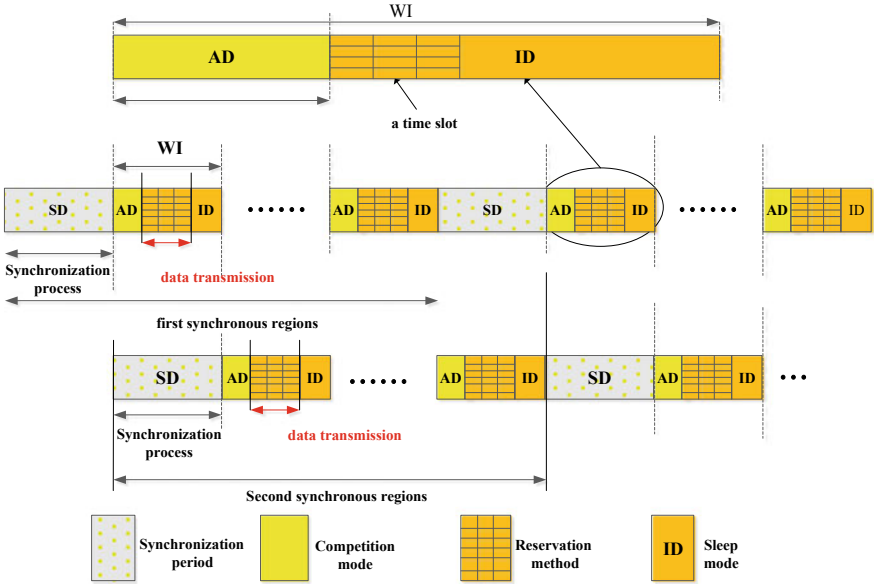


Fig. 3. The second stage of the synchronization process: the synchronization stage

Once the synchronization error is updated, it may exceed a given threshold and trigger the synchronization process again. After the first node is synchronized with the root node, the reference nodes are selected according to the energy and node degree of the sub-nodes, and the synchronization process is repeated until all the nodes in the network arrive at the synchronization.

3 Analysis Index of High-Performance Synchronous Energy-Saving Algorithm

The analysis index of high-performance synchronous energy-saving algorithm is analyzed qualitatively from the point of view of message transmission energy and energy consumption. Figure 4 shows the energy consumption diagram of the intermediate node. In the level discovery phase, compared to the TPSN, the parent node accepts more child level discovery packets from the child nodes, but the level discovery is only performed in the network initialization, so the increased message transmission volume is relatively small compared to the synchronization.

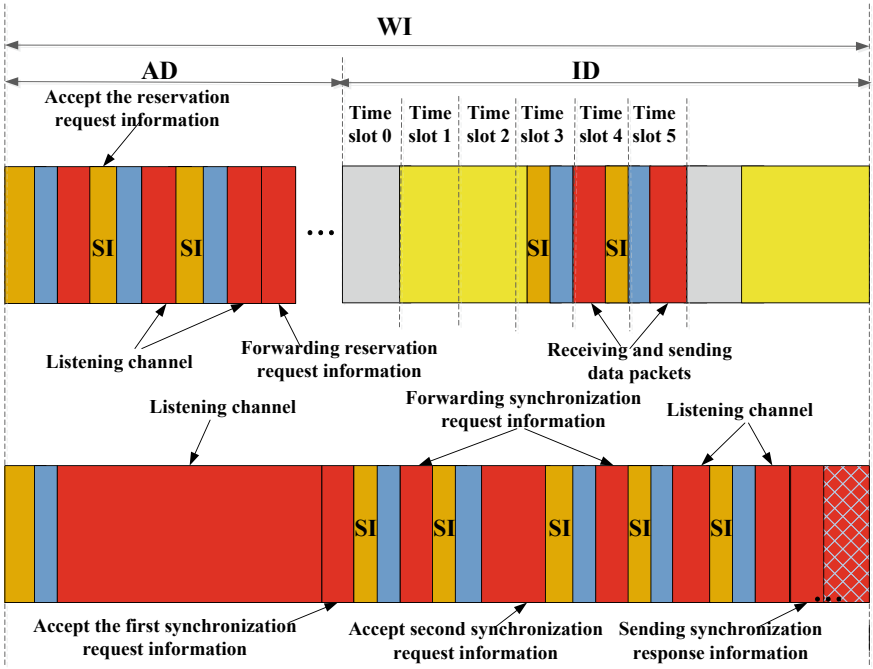


Fig. 4. The energy consumption diagram of the intermediate node

In synchronization, the reference node is synchronized to the parent node for TPSN each time. If the parent node has a lot of sub-nodes, the amount of information reduced will be very large, which greatly reduces the energy consumed by the intermediate

nodes because of the assimilation. In addition, a reference node is selected by considering the energy and node degree of the sub-nodes, which ensures the balance of energy consumption. It greatly reduces the vicious circle of hierarchical discovery initiated by the intermediate node because of the unsuccessful synchronization of the sub-nodes caused by the energy depletion, which greatly prolongs the lifetime of the network. In some large multi-hop networks, the lifetime of the network is mainly affected by the intermediate nodes.

4 Simulation Results

In order to verify the performance of high-performance synchronous energy-saving algorithm, this paper simulates and analyzes the performance of high-performance synchronous energy-saving algorithm in wireless sensor networks based on NS-2 (Network Simulator Version 2) simulation environment, and compares it with IEEE 802.15.5 in traditional synchronous energy-saving mode [9]. The simulation environment has set up a 625 (25 * 25) m^2 regular grid layout, randomly deployed 50 nodes, and the distance between two successive neighbor nodes is 30 m. In this topology, the grid coordinator is placed in the upper left corner of the grid (node 0), and the other nodes are randomly distributed. The node frequency drift obeys normal distribution, the synchronization period is set to 50 s, and the simulation time is 800 s, data transfer rate is 250 kbps.

In order to assess the performance of high-performance synchronous energy-saving algorithms (HPSESA), we choose different indicators such as throughput (bandwidth utilization), delay (average delay), jitter (deviating the average message delay), message transfer rate, and simulation results as shown in Fig. 5.

The results of HPSESA simulation and analysis are shown in Fig. 5, and compared with the traditional synchronous energy-saving algorithms (SES). As shown in Fig. 5a, the throughput of HPSESA is higher than that of the traditional SES based on IEEE 802.15.5. Our algorithm avoids special synchronous slot and frees more time for data transmission. Compared with SES, the HPSESA algorithm significantly reduces the average delay of messages, as illustrated by Fig. 5b. It can be concluded that the coincidence of the synchronization process message and the probability of data transmission operation time are very low in the reply link. Therefore, information flows continuously between the source transmitter and receiver, and the sleep state of nodes is delayed at the lowest level. On the other hand, because the synchronization process does not increase the transmitted messages, the jitter of the high-performance synchronous energy-saving algorithm decreases dramatically, which can be seen from Fig. 5c. It can also be seen from Fig. 5d that HPSESA delivers data messages with a 100% success rate for messaging, thus providing more data messages to avoid retransmission.

Figure 6 compares the energy consumption of nodes per bit in HPSESA and SES algorithm. Compared with SES algorithm, HPSESA has less energy consumption and can maintain strict synchronization. During the duration of the activity, since HPESEA has better synchronization accuracy and can provide higher information, the time

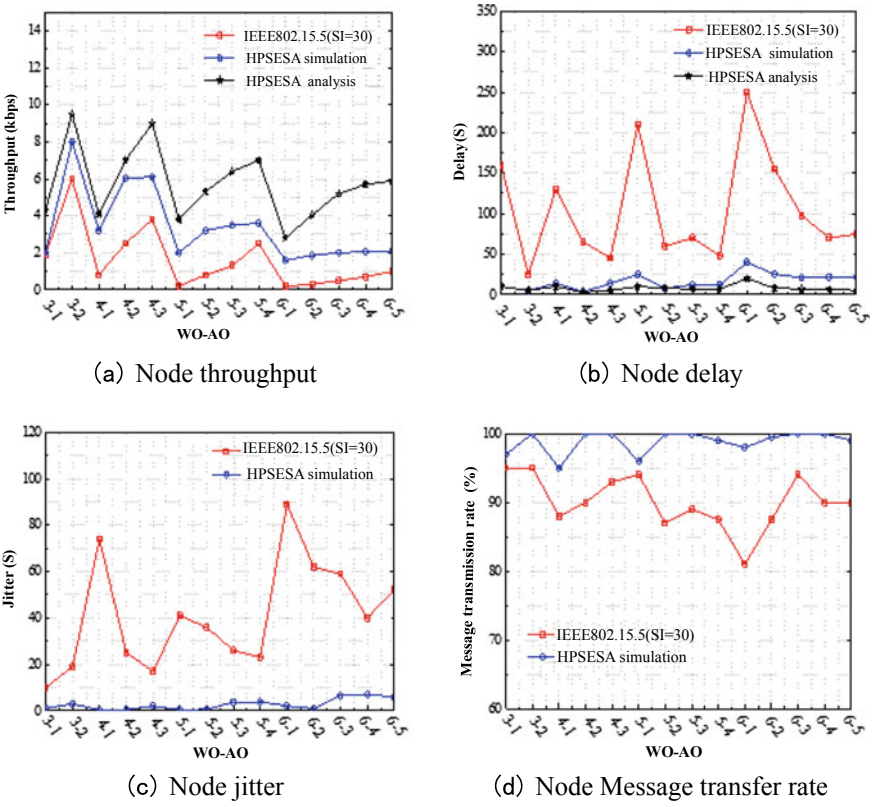


Fig. 5. HPSESA simulation and analysis results

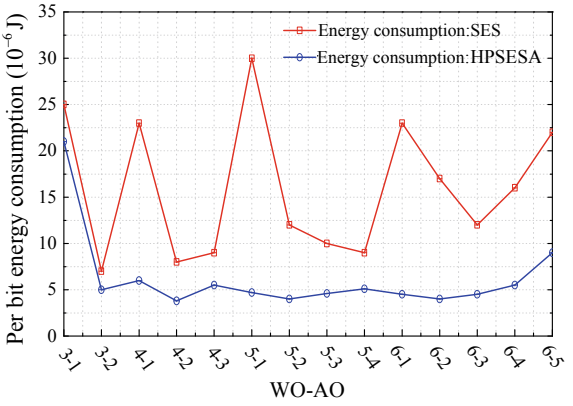


Fig. 6. Energy consumption of SES and HPSESA algorithms

needed to send the same data is reduced. At the same time, since the energy of each node of HPSESA is well balanced, it can greatly extend the life of the whole network, which is HPSESA's biggest advantage.

5 Conclusion

In this paper, we propose a new high-performance synchronization energy-saving algorithm for wireless sensor networks that enhance the existing time synchronization algorithm. The algorithm has shorter synchronization convergence time than FTSP and uses less energy than other time synchronization algorithm. Considering the energy constraints of wireless sensor networks, the algorithm can not only achieve better synchronization accuracy but also prolong the network lifetime from energy saving and energy consumption of balanced nodes. The purpose of our future research is to improve the synchronization accuracy of high-performance synchronization energy-saving algorithm.

References

1. Tsitsiklis, J.: Design of optimal energy aware protocols for wireless sensor networks. In: Proceedings of IEEE International Conference on Wireless Days, pp. 1–8 (2012)
2. Ping, S.: Delay Measurement Time Synchronization for Wireless Sensor Networks, Intel Research Center: IRB-TR-03-013 (2003)
3. Dai, H., Han, R.: TSync a lightweight bidirectional time synchronization service for wireless sensor networks. ACM SIGMOB ILEMobile Comput. Commun. Rev. **8**(1), 1–4 (2004)
4. Shah, R., Rabaey, J.: Energy aware routing for low energy in wireless sensor networks. J. Wirel. Commun. **2**(12), 35–40 (2012)
5. Franceschelli, M., Giua, A., Seatzu, C.: The capacity of energy-constrained mobile networks with wireless power transfer. IEEE Trans. Photonics Technol. Lett. **26**(8), 781–784 (2014)
6. Zhang, W.K., Liu, P., et al.: Research on the algorithm of energy-aware time synchronization for wireless sensor networks. In: International Conference on, pp. 3545–3548 (2011)
7. Wu, S.C., Rabbat, M.G.: Distributed consensus with limited communication data rate. IEEE Trans. Signal Process. **61**(16), 3959–3971 (2013)
8. Torkestani, J.A.: Internet time synchronization: the network time protocol. J. Netw. Comput. Appl. **23**(9), 1633–1635 (2012)
9. The Network Simulator ns-2[EB/OL]. http://www.isi.edu/nsnam/ns/tutorial/index.html. [2006-05-15]/[2015-04-04]

Multi-user Directional Modulation Based on Polarization Transmission

Zhiyong Chen(✉), Aijun Liu, Xiaohu Liang, and Xuan Yi

Army Engineering University of PLA Nanjing, Nanjing 210007, China
3120104308@zju.edu.cn

Abstract. In this paper, we proposed a novel multi-users directional polarization modulation (MUDPM) based on directional modulation and polarization modulation. The scheme can achieve security multi-users polarization modulation. The feasibility and security of MUDPM are evaluated in this paper. Moreover, we introduce a secure and precise transmission scheme. The scheme provides location-specific secure communication to legitimate users and the BER spatial distributions of the proposed scheme are analyzed via simulation.

Keywords: Directional modulation · Polarization modulation · Physical layer security

1 Introduction

Wireless communication technology plays an extremely important role in civil and military applications. Because of the broadcast nature of the wireless channel, the wireless information transmission leads to confidentiality worries. Security solutions have been proposed at the physical layer to improve the security of wireless transmissions. Directional modulation (DM) is a promising physical layer security technology that can be realized by using antennas array [1–7]. DM technology makes the modulation realization depend on the direction of signal. The constellation patterns can be received exactly in the predetermined communication direction and it will be distorted in the unintended direction. In [1,2], the DM technology using the phased array antenna is realized in the radio-frequency domain. The high-precision RF phase shifter that needs to switch at the information symbol rate is usually not easy to realize from the practical point of view. In [3], a more simple and accurate amplitude–phase-controlled digital DM structure is proposed. Afterward, [4] develops the constrained far-field radiation pattern method, and the method for separating the far-field radiation pattern is given in [5]. In [6], the authors propose the orthogonal vector method to generate DM signal. They define the DM structure as two categories. The first category is "Static" DM structure, where the generated antenna patterns are the same over time, and it is easy to be deciphered if eavesdropper gets the randomization pattern along his direction via statistical method. The second category is called "Dynamic" DM structure, where the same symbol can be

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 678–685, 2019.
https://doi.org/10.1007/978-981-13-6264-4_80

transmitted using a random antenna pattern each time, which makes it hard to be deciphered. In [7], a secure multiple access DM scheme is proposed, which exploits DM and the multi-path structure of the channel to achieve multi-user physical layer security transmission.

The traditional modulation modulates information through amplitude, phase, and frequency of electromagnetic waves, such as amplitude modulation, phase modulation, frequency modulation, and amplitude–phase modulation. With the application of orthogonal dual-polarized antennas, the experts pay more attention to the research of polarization signal processing. Polarization can also be used as a way to carry information. Inspired by the polarization modulation in optical fiber communication, Wei Dong [8] proposed polarization characteristics using electromagnetic wave carrying information. And in order to further improve the utilization rate of the electromagnetic wave, Wei Dong [9] proposed polarization amplitude–phase modulation. However, the above DM works do not apply to the polarization transmission technology. Polarization modulation (PM) is an important polarization signal processing technology in polarization transmission applications and its security transmission is a problem that needs to be solved.

In this paper, we design a novel multi-user directional polarization modulation (MUDPM) based on directional modulation and polarization modulation. First, we design the system model and the signal structure. Second, we verify the feasibility and security of MUDPM scheme. Third, we design a kind of secure and precise wireless communication system based on MUDPM, and it can ensure that the useful information can be received accurately in a small neighborhood near the desired position. Lastly, we evaluate the system performance via simulation and analysis.

2 Multi-user Directional Modulation System Design

We consider a broadcast channel with a single base station and L users. Every user has its own independent information string $\{I_{ij}\}_{j=1}^{n}$ and transmission angle with respect to the base station, θ_i, where i = 1,2,. . . ,L. Different users share the same resources of time slots, frequency bands, or codes simultaneously. The base station uses a linear antenna array, with N pairs of orthogonal dual-polarized antennas spacing of one-half wavelength, for transmission. Receivers use a pairs of orthogonal dual-polarized antennas. In this paper, we assume that the antennas with different polarizations are almost completely orthogonal and we do not consider the interference between different polarization states.

For a 1D N-element single-polarized isotropic antennas array with one-half wavelength spacing, the array steering vector $\boldsymbol{h}(\boldsymbol{\theta})$ is

$$\boldsymbol{h}(\theta) = [e^{j\pi(1-(N+1)/2)\cos\theta}, e^{j\pi(2-(N+1)/2)\cos\theta}, \ldots, e^{j\pi(N-(N+1)/2)\cos\theta}] \quad (1)$$

So the array steering matric $H(\theta)$ is

$$H(\theta) = \boldsymbol{h}(\theta) \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{2}$$

where $\otimes$ represent Kronecker product.

In the MUDPM system, transmitter modulates messages through amplitude–phase modulation and polarization modulation. In polarization modulation, we can express the polarization state $\boldsymbol{P}$ with Jones vector as

$$\boldsymbol{P} = \begin{bmatrix} \cos\delta \\ \sin\delta e^{j\varphi} \end{bmatrix} \tag{3}$$

where (δ, φ) $(\delta \in [0, \pi/2], \varphi \in [0, 2\pi])$ is called the polarization phase descriptor.

For the ith user, the information string $\{I_{ij}\}_{j=1}^{n}$ is divided into two parts: $\{I_{ik}\}_{k=1}^{K}$ and $\{I_{im}\}_{m=1}^{M}$. After that, $\{I_{ik}\}_{k=1}^{K}$ is carried on to $\{S_{ik}\}_{k=1}^{K}$ by the amplitude–phase modulation and $\{I_{im}\}_{m=1}^{M}$ is modulated to $\{cos\delta_{im}\}_{m=1}^{M}$ and $\{sin\delta_{im}e^{j\varphi_{im}}\}_{m=1}^{M}$ by the polarization modulation. Then, the modulated information data stream $x_{iH}(km) = S_{ik} \cdot \cos\delta_{im}$, and $x_{iV}(km) = S_{ik} \cdot \sin\delta_{im}e^{j\varphi_{im}}$ are input into digital MUDM and signal vectors are output from digital MUDM. Lastly, signal vector are input to the corresponding antennas via RF and PA.

The signal vector in the array is

$$\boldsymbol{D}_{km} = [d_{H,1}, d_{V,1}, d_{H,2}, d_{V,2}, \ldots d_{H,N}, d_{V,N}]^T \tag{4}$$

$d_{X,i}$ denotes signal scalar in ith X-polarized transmitting antenna.

$$\boldsymbol{D}_{km} = \boldsymbol{D}_{Hkm} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \boldsymbol{D}_{Vkm} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{5}$$

$\boldsymbol{D}_{Hkm}$ and $\boldsymbol{D}_{Vkm}$ are the signal vectors in horizontally polarized antennas and vertically polarized antennas, respectively, which meet

$$\boldsymbol{E}_{km}(\theta_i) = H(\theta_i) \cdot \boldsymbol{D}_{km} = \begin{bmatrix} \boldsymbol{h}(\theta_i) \cdot \boldsymbol{D}_{Hkm} \\ \boldsymbol{h}(\theta_i) \cdot \boldsymbol{D}_{Vkm} \end{bmatrix} = \begin{bmatrix} S_{ik} \cdot \cos\delta_{im} \\ S_{ik} \cdot \sin\delta_{im}e^{j\varphi_{im}} \end{bmatrix} \tag{6}$$

So, we can get the signal vectors $\boldsymbol{D}_{Hkm}$ and $\boldsymbol{D}_{Vkm}$ are

$$\begin{aligned} \boldsymbol{D}_{Hkm} &= \boldsymbol{H}^*(\boldsymbol{H}\boldsymbol{H}^*)^{-1}\boldsymbol{x}_H(km) \\ \boldsymbol{D}_{Vkm} &= \boldsymbol{H}^*(\boldsymbol{H}\boldsymbol{H}^*)^{-1}\boldsymbol{x}_V(km) \end{aligned} \tag{7}$$

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{h}(\theta_1) \\ \boldsymbol{h}(\theta_2) \\ \vdots \\ \boldsymbol{h}(\theta_L) \end{bmatrix} \tag{8}$$

3 Security Analysis

The received signal in communication direction θ for any eavesdropper is

$$\boldsymbol{E}_{km}(\theta) = H(\theta) \cdot \boldsymbol{D}(km) = \begin{bmatrix} \boldsymbol{h}(\theta) \cdot \boldsymbol{H}^*(\boldsymbol{H}\boldsymbol{H}^*)^{-1}\boldsymbol{x}_H(km) \\ \boldsymbol{h}(\theta) \cdot \boldsymbol{H}^*(\boldsymbol{H}\boldsymbol{H}^*)^{-1}\boldsymbol{x}_V(km) \end{bmatrix} = \begin{bmatrix} \boldsymbol{v} \cdot \boldsymbol{x}_H(km) \\ \boldsymbol{v} \cdot \boldsymbol{x}_V(km) \end{bmatrix} \tag{9}$$

We assume that eavesdropper want to get the information of the lth user. The polarization state received by eavesdropper is

$$\boldsymbol{P}_e = \begin{bmatrix} \cos\delta_e \\ \sin\delta_e e^{j\varphi_e} \end{bmatrix} \tag{10}$$

$$\delta_e(\theta) = arctan\left(\frac{\boldsymbol{v} \cdot \boldsymbol{x}_V(km)}{\boldsymbol{v} \cdot \boldsymbol{x}_H(km)}\right) = arctan\left(\frac{v_l \cdot x_{lV}(km) + \sum_{j\neq l} v_j \cdot x_{jV}(km)}{v_l \cdot x_{lH}(km) + \sum_{j\neq l} v_j \cdot x_{jH}(km)}\right) \tag{11}$$

$$\begin{aligned} \varphi_e(\theta) &= \varphi(\boldsymbol{v} \cdot \boldsymbol{x}_V(km)) - \varphi(\boldsymbol{v} \cdot \boldsymbol{x}_H(km)) \\ &= \varphi(v_l \cdot x_{lV}(km) + \sum_{j\neq l} v_j \cdot x_{jV}(km)) - \varphi(v_l \cdot x_{lH}(km) + \sum_{j\neq l} v_j \cdot x_{jH}(km)) \end{aligned} \tag{12}$$

We define $\boldsymbol{v} \cdot \boldsymbol{x}_V(km)$ and $\boldsymbol{v} \cdot \boldsymbol{x}_H(km)$ as useful message and define $\sum_{j\neq l} v_j \cdot x_{jV}(km)$ and $\sum_{j\neq l} v_j \cdot x_{jH}(km)$ as interference. So, we can get the signal-to-interference-plus-noise-ratio (SINR) in every communication directionθ as shown in Fig. 1. In the shown result, $N = 10$, $L = 4$, $\theta_l = 80°$ and we assume signal-to-noise ratio (SNR) is 10. The figure shows extremely low SINR outside the main lobe, which indicates that the data obtained in the unintended communication direction cannot be detected reliably.

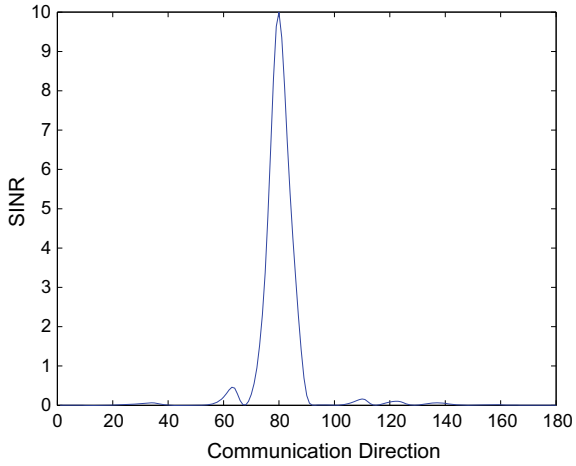


Fig. 1. SINR of received signal in different communication direction.

When we fix the $x_H(km)$ and $x_V(km)$ as contents, i.e., for any time index k and m, $x_H(km)$ and $x_V(km)$ is same, we can classify this MUDPM as a kind of static directional modulation. When $x_{jH}(km)$ and $x_{jV}(km)$ change with the time index k and m, the interference will change with k and m, then we can classify this MUDPM as a kind of dynamic directional modulation.

4 Secure and Precise Wireless Transmission

For intended receivers and eavesdropper with the same direction angle, the traditional DM technology cannot continue to ensure the security of information transmission. The secure and precise wireless communication based on MUDPM is proposed to achieve that the useful information can be received accurately in a small neighborhood near the desired position, while in other areas it is impossible to receive the useful information accurately for eavesdropper.

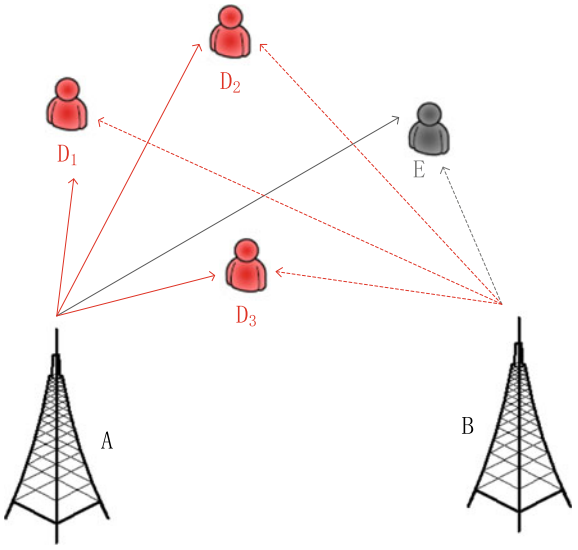


Fig. 2. secure and precise wireless transmission system model

As shown in Fig. 2, we have designed a secure and precise wireless communication system based on MUDPM, in which Ds are desired receiver, E is a eavesdropper, A and B are two independent base stations, two base stations send signals of orthogonal polarization states through MUDPM, so that two polarization state signals can be correctly received at a small neighborhood near the desired receiver D.

The base station A uses a vertical polarized linear antenna array to transmit the vertical polarization signal vector.

$$\boldsymbol{D}_{Vkm} = \boldsymbol{H}_A^*(\boldsymbol{H}_A\boldsymbol{H}_A^*)^{-1}\boldsymbol{x}_V(km) \tag{13}$$

The base station B uses a horizontal polarized linear antenna array to transmit the horizontal polarization signal vector.

$$\boldsymbol{D}_{Hkm} = \boldsymbol{H}_B^*(\boldsymbol{H_B}\boldsymbol{H}_B^*)^{-1}\boldsymbol{x}_H(km) \tag{14}$$

The received signal is

$$\boldsymbol{E}_{km} = \begin{bmatrix} \boldsymbol{h}(\theta_B)\cdot\boldsymbol{H}_B^*(\boldsymbol{H_B}\boldsymbol{H}_B^*)^{-1}\boldsymbol{x}_H(km) \\ \boldsymbol{h}(\theta_A)\cdot\boldsymbol{H}_A^*(\boldsymbol{H_A}\boldsymbol{H}_A^*)^{-1}\boldsymbol{x}_V(km) \end{bmatrix} \tag{15}$$

So, only when the receiver is at a small neighborhood near the desired receiver, the received signal do not get affected by the artificial interference.

5 Simulation Result and Analysis

To evaluate the dynamic DPM scheme proposed in this paper, its BER simulation is carried out under the following prerequisites: First, 4PM and QPSK are used as modulations in MUDPM system. Second, A 1D eight pairs of dual-polarized antennas array with one half wavelength spacing is used in transmitter, and the number of users is eight. Last, the SNR of the AWGN channel is 10 dB in this system.

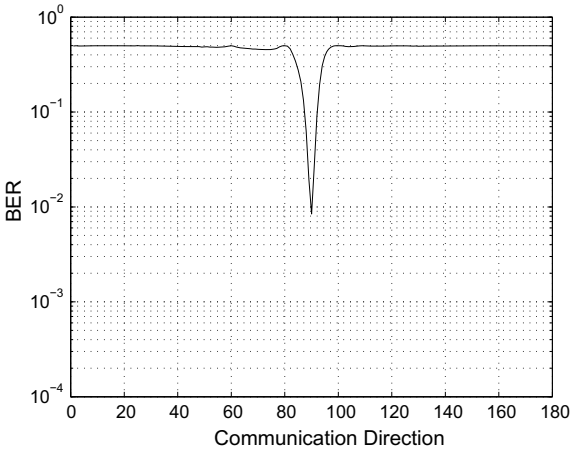


Fig. 3. BER in different communication direction.

Figure 3 shows the BER performance of MUDPM system for intended communication direction 90°. It can be seen that the signal can be demodulated correctly by the receiver in the desired communication direction and cannot be demodulated in undesired communication direction. So, the physical layer security of multi-user polarization modulation has been achieved in MUDPM.

The secure and precise transmission scheme requires the signal to be transmitted from two geographically separated base stations to provide security so that, along each direction of transmission, the data is secure. Here, we simulate a 200 × 200 area. The number of users and or orthogonal dual-polarized antenna pairs is 8($N = L = 8$). Using QPSK and 4PM modulation scheme, the base stations are in (200,0) and (0,0), and the noiseless BER performance for one of the users which located in (48,92) is shown in Fig. 4.

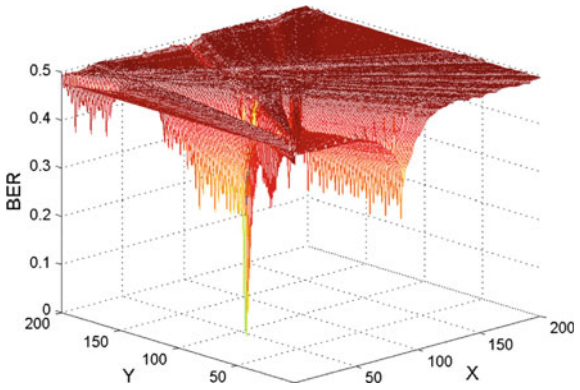


Fig. 4. BER spatial distributions of secure and precise wireless transmission system based on MUDPM.

It can be seen from the Fig. 4, for the secure and precise wireless communication system based on MUDPM the receiver can receive the signal correctly only in a certain neighborhood of the expected receiver.

6 Conclusion

In this paper, a novel MUDPM scheme is proposed for realizing security multi-user polarization modulation. The feasibility and security of MUDPM scheme is verified. We design a kind of secure and precise wireless communication system based on MUDPM, and it can ensure that the useful information can be received accurately in a small neighborhood near the desired position. The MUDPM scheme extends the application range of directional modulation and polarization modulation technique.

References

1. Daly, M.P., Bernhard, J.T.: Directional modulation technique for phased arrays. IEEE Trans. Antennas Propag. **57**(9), 2633–2640 (2009)
2. Daly, M.P., Daly, E.L., Bernhard, J.T.: Demonstration of directional modulation using a phased array. IEEE Trans. Antennas Propag. **58**(5), 1545–1550 (2010)

3. Snow, T., Chappell, W.J.: Directional modulation using independent waveforms at each element of a phased array radar. In: IEEE MTT-S International Microwave Symposium Digest, pp. 1–4 (2013)
4. Ding, Y., Fusco, V.F.: Directional modulation transmitter radiation pattern considerations. IET Microw.S, Antennas Propag. **7**(15), 1201–1206 (2013)
5. Ding, Y., Fusco, V.F.: Directional modulation far-field pattern separation synthesis approach. IET Microw.S, Antennas Propag. **9**(1), 41–48 (2014)
6. Ding, Y., Fusco, V.F.: A vector approach for the analysis and synthesis of directional modulation transmitters. IEEE Trans. Antennas Propag. **62**(1), 361–370 (2014)
7. Hafez, M., Yusuf, M., Khattab, T., et. al.: secure spatial multiple access using directional modulation. IEEE Trans. Wirel. Commun. **17**(1), 563–573 (2018)
8. Dong, W., Feng, C., Guo, C., et al.: A power amplifier energy efficient polarization modulation scheme based on the optimal pre-compensation. Commun. Lett. IEEE **17**(3), 513–516 (2013)
9. Wei, D., Feng, C., Guo, C., et al.: An optimal pre-compensation based joint polarization-amplitude-phase modulation scheme for the power amplifier energy efficiency improvement. In: Proceedings of IEEE International Conference on Communication(ICC) (2013)

Improving Convergence of IEEE 1588v2 Protocol Using Clock Servo System

Weiwen Gu[1], Yue Yu[1], Heqin Tong[2], Qin Gao[1], Zhijian Zhang[1], and Jian Wang[1(✉)]

[1] School of Electronic Science and Engineering, Nanjing University, 163 Xianlin Road, Nanjing 210023, Jiangsu, China
18351881316@163.com, 1419754775@qq.com, {qingao,njuzzj,wangjnju}@nju.edu.cn
[2] State Key Laboratory of Smart Grid Protection and Control, NARI Group Corporation, 19 Chengxin Road, Nanjing 211000, Jiangsu, China
tongheqin@sgepri.sgcc.com.cn

Abstract. Precise time synchronization is widely used in several domains such as Internet of Things (IoT), communication industry, power system and so on. Among all-time synchronization protocols, IEEE 1588v2 Protocol has received extensive attention, due to its submicrosecond accuracy and high-cost performance. However, it is not great enough in the convergence performance of synchronization in software-only system. Therefore, the aim of this paper is to speed up the convergence by control the jitter in software-only system. There are several methods proposed to reduce the jitter brought by soft timestamp and nonreal-time operating system. We compare and analyze various methods and choose clock servo system to experiment. The results of experiments verify that the clock servo system reduces the convergence time and improves the synchronization accuracy to some extent.

Keywords: IEEE 1588v2 · Clock servo system · Jitter · Convergence · Accuracy · Offset from master

1 Introduction

Time synchronization is widely considered for the measurements of the network delay, which is used to avoid message transmission errors caused by more and more clock nodes in a distributed system [1]. Compared to the widely used NTP protocol, IEEE 1588v2 Protocol can achieve higher accuracy which can meet submicrosecond synchronization requirements. Furthermore, it takes up a small amount of CPU resources which can run in embedded system [2]. Therefore, it gets the favor of the internet of things and other industries like telecommunication, smart power grid systems and so on [3].

In order to study the influence factors, Giorgi analyzes effects of the instability of local clocks, the rate, and the accuracy of the results and their relationships

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 686–693, 2019.
https://doi.org/10.1007/978-981-13-6264-4_81

[4]. The researches of IEEE 1588v2 Protocol focused on improving synchronization performance in different scenarios [5–8]. However, the timestamp in most of them is provided by hardware clock, which is in the physical layer and accurate enough. A software-only system will bring uncertain jitter, which will extend convergence time and reduce synchronization accuracy.

To solve the problem brought by jitter, there are serval solutions presented. Giorgi models software timestamp errors as the sum of Bernoulli and Gaussian random variables and the nondeterministic features in the abovementioned Linux system can be simulated by Bernoulli variables. If this disturbance can be controlled, jitter can be minimized [9]. Li Qin proposed an easy-to-operate PTN solution which eliminates unpredictable risks by establishing a highly controllable network with clock and time separation [10]. However, it requires specific product equipment. Correll K. proposed an architecture based on wired synchronization clock servo system [11]. In this paper, we analyze the source of jitter and compare the solutions. By experiments, the results show that the clock servo system can greatly improve the performance of software-only system.

This paper is organized as follows. In Sect. 2, we overview the mechanism of IEEE 1588v2 Protocol. Problems in software-only systems are analyzed in Sect. 3. In Sect. 4, we describe the clock servo system for improving clock synchronization performance. The results of the experiments are presented and discussed in Sect. 5. In Sect. 6, we draw the conclusions.

2 IEEE 1588v2 Mechanism

In this section, we introduce the time synchronization mechanism of precision time protocol (PTP) version two protocol, IEEE 1588v2. A PTP system is constituted by one or more PTP domains and a PTP domain is constituted by PTP clocks and PTP communication channels. Most of the clocks are ordinary clock and boundary clock. There is one clock port in the ordinary clock and several clock ports in boundary clock. Therefore, boundary clocks can connect to different PTP communication channels. We use best master clock (BMC) algorithm to compare the accuracy of every PTP clock quantitatively. Clock messages are received through PTP port into different data sets and these binary dates are used to calculate the most accurate clock as master clock. It is worth noting that clock state in the whole system is not immutable. If number or accuracy of clock is changed, the clock state will also change accordingly.

Local clock synchronization algorithm is used in slave clocks to synchronize with the master clock. After receiving time messages from the master clock, the slave clock calculates the offset between the local clock and the master clock and modifies his own local clock. The specific synchronization process is shown in Fig. 1. The master clock regularly sends a *Sync* message to the slave clock containing the sent time *TM1*. The time of *Sync* message reaching the slave clock is recorded as the received time *TS1*. The slave clock then transmits a *Delay_Req* message containing the sent time *TS2* to the master clock. Finally, the master clock records the reception time *TM2* and sends *Delay_Resp* message containing it to the slave clock [12].

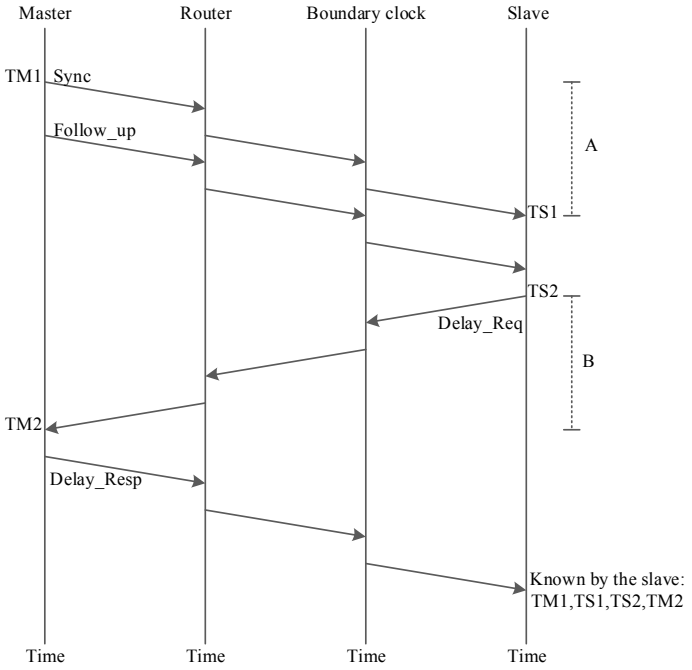


Fig. 1. Local clock synchronization algorithm

The one-way delay and the offset from master at the slave clock is approximated by:

$$D = one - waydelay = (A + B)/2, \tag{1}$$

$$\theta = offsetfrommaster = (A - B)/2. \tag{2}$$

where A = TS1 − TM1 and B = TM2 − TS2 correspond to the the master–slave delays and the slave–master delays respectively. To correct its clock, the slave clock adjusts its local time t to t − θ. We think that the messages experience symmetrical delays, which will influence the synchronization accuracy.

3 Proposed Method Using Clock Servo System

3.1 Problem in Software-only System

Convergence performance of IEEE 1588v2 Protocol affects the speed of synchronization. If the convergence time is too long, it will lead to cumulative error which has a great influence on clock accuracy. Of course, there are several factors affecting the convergence performance, such as the location of the timestamp, system category, communication channel congestion and so on.

If the system is software-only, the timestamp can only be in the higher network layer. Compared to the hardware timestamp in the physical layer, it introduces indeterminable latencies in timestamp. In a nonreal-time operating system,

like Linux system, the CPU cannot be preempted. Then the wait time of the task cannot be determined, which will generate jitter in the delay estimation. In addition, operating system latency caused by context switching may affect the accuracy of the timestamp, while cross-traffic and interaction with PTP packets may also affect packet delay. All of these jitters will eventually result in a decrease in synchronization accuracy and longer convergence time.

4 Architecture of IEEE 1588v2 with Clock Servo System

The IEEE 1588v2 Protocol does not define the clock servo system, nor does it define how to perform offset adjustment to achieve master–slave clock synchronization. As mentioned in Sect. 3.1, due to jitter in a software-only system, it is necessary to use a clock servo system. The architecture of IEEE 1588v2 Protocol is shown in Fig. 2. Different from the conventional methods, the timestamp is in the network layer, and the clock servo system is an additional process. They are all marked in red.

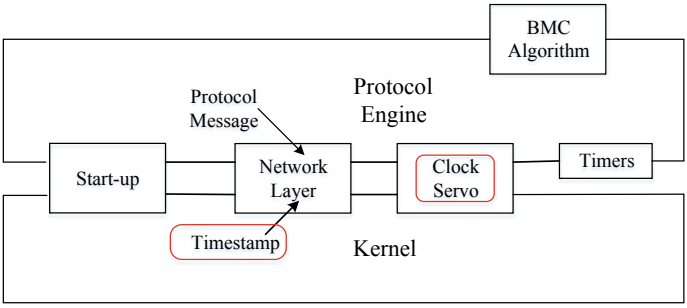


Fig. 2. Architecture of IEEE 1588v2 with clock servo system

4.1 Clock Servo System

The controller is the core of the entire clock servo system, which makes the offset and delay converge. The most commonly used is a kind of liner controller: the PI controller, consisting of proportional regulation and integral regulation. Proportional regulation is used to reduce system deviation and track correction input. If the proportional adjustment parameter increases, the adjustment speed becomes faster and the error becomes smaller. Integral regulation is used to eliminate, track and correct the system steady-state error. The effect of integral regulation changes reversely with the integral constant [13].

Among the information of input, the frequency of effective information is close to zero, and the frequency of the noise is relatively high. Therefore, we can attenuate noise in the input signal to the maximum extent through filtering out the input of continuous noise and impulse noise by the low-pass filter, which

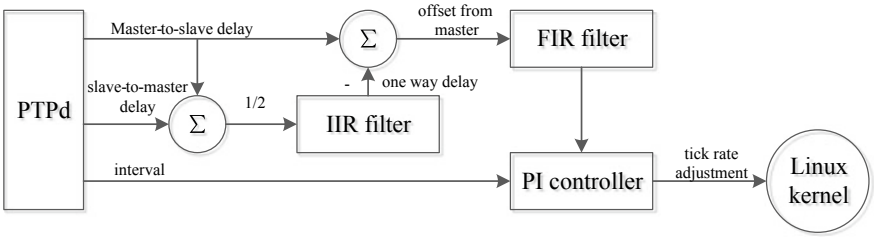


Fig. 3. Clock servo system proposed by Correll K

reduces jitter introduced into the controller. Thus, the conventional clock servo systems (Fig. 3) consist of low-pass filters in addition to the PI controller.

The infinite impulse response (IIR) low-pass filter is used to filter out one-way delay. The protocol samples master–slave delays periodically and slave–master delays are occasionally collected to ensure that the master clock is the most accurate clock in the system. The formula of the IIR filter is shown in Eq. 3.

$$s * y[n] - (s-1) * y[n-1] = \frac{x[n]}{2} + \frac{x[n-1]}{2} \tag{3}$$

where *x[n]* represents one-way delayed signals, *y[n]* represents the filtered output, and s is the control variable of the IIR low-pass filter. Although the variable s directly influences the cutoff frequency, the cutoff frequency of the filter cannot be unrestrictedly reduced because it is at the expense of extend convergence time.

Compared to the IIR low-pass filter, the phase of FIR low-pass filter is accurately linear. The filter delay of the FIR low-pass filter in the clock servo system is one sampling period, which is the lowest delay of the filter. In order to avoid undeterminable delay errors, we use FIR low-pass filters to filter out the offset from master. The formula of the FIR low-pass filter is shown in Eq. 4.

$$y[n] = \frac{x[n]}{2} + \frac{x[n-1]}{2} \tag{4}$$

where *x[n]* represents the value of offset from master and *y[n]* represents the output value of the filter.

5 Results

We conducted the experiment for both the proposed (with clock servo system) and conventional (without clock servo system) methods in order to verify the improvement of synchronization accuracy and convergence performance compared to the conventional method. We use two PCs which are connected to the same router by cables to form a small LAN for testing, Fig. 4 is the result.

For the conventional method, initial value of the offset from master is about 100 ms. It takes 8 min to converge to 1 ms, and 15 min to converge to 10 us, shown

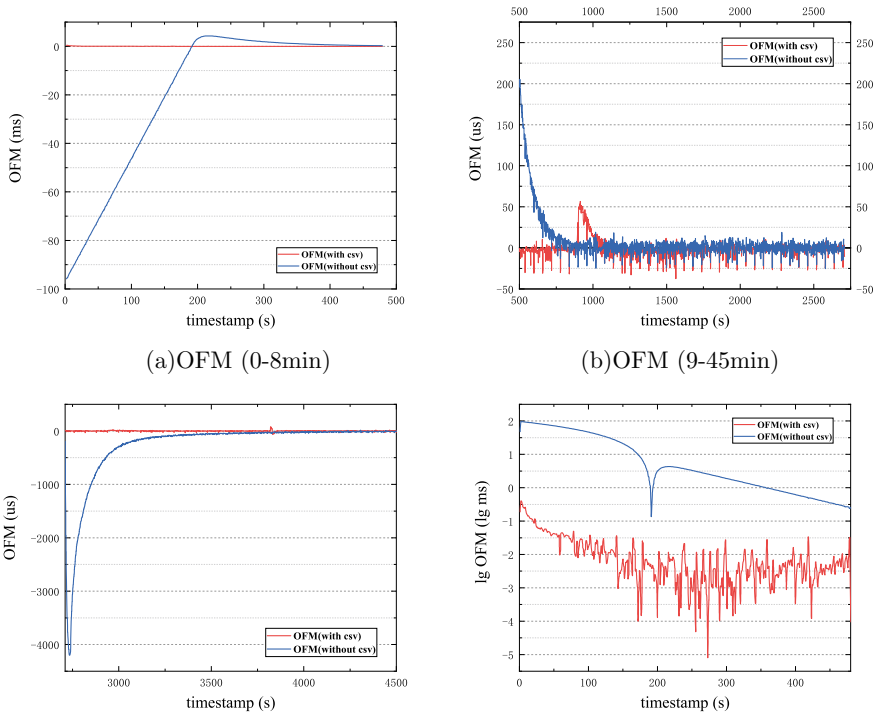


Fig. 4. Comparing OFM with csv and without csv in different time scales (OFM and csv respectively represent offset from master and clock servo system)

in Fig. 4a. Correspondingly in Fig. 4c, when there is a transient instability in the network, it takes a long time to reconverge to 10 us. On the other hand, initial value of the offset from master is lower than 1 ms for the proposed method. It converges to 10 us after 3 min and maintains a small amplitude fluctuation around 10 us in a long term. Figure 4b shows that the offset from master reaches about 50 us, due to temporary instability of the network at the fifteenth minute. But it can re-converge to 10 us within about 2 min. Compared with conventional method, the convergence speed is faster and the value of the offset from master is more stable when there is a transient instability in the network.

Initially, the response characteristic of PI controller dominates the convergence, and the control variable s of IIR filter during this period is low. After the system is stable, the control variable s of IIR filter which dominates in the convergence increases continuously [14]. Figure 4d is the logarithmic change of the offset from master within 8 min. It is clearly that the initial value in proposed method is less than zero, which means that the offset from master is less than 1 ms. In conventional method, it needs 8 min converge to achieve the same level. Figure 5 is the logarithmic change of the offset from master within 45 min.

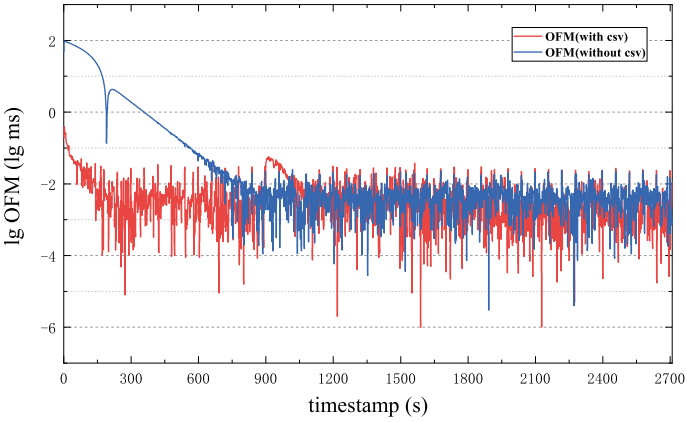


Fig. 5. Logarithm of OFM within 45 min

The values in both methods are less than −2 after 15 min, which means that both methods converge to 10 us within 15 min. However, the proportion of value that less than −3 is twenty-one percent, three times of the conventional method. It is closer to the theoretical value of 0.1 us of the PTP protocol.

6 Conclusion

In this paper, we test a PTP system incorporating clock servo which improves convergence speed and synchronization accuracy. The conventional IEEE 1588v2 Protocol uses hardware clock sources. In the case of software-only system, the clock source is in the network layer. The clock servo system filters out the noise and controls the clock signal stably. Therefore, this method solves the jitter problems brought by network layer and nonreal-time operating system. It eventually speeds up convergence and improves accuracy. For the improvement of the clock servo system, we can consider the introduction of nonlinear filter components and nonlinear controllers to achieve better synchronization accuracy. We can also improve the precision and convergence speed by improving the clock servo system in other communication channels such as wireless, SDH and so on.

Acknowledgments. This work was supported by State Key Laboratory of Smart Grid Protection and Control of NARI Group Corporation.

References

1. Shin, M., Park, M., Deockgil, O., Kim, B., Lee, J.: Clock Synchronization for One-Way Delay Measurement: A Survey. Springer, Berlin, Heidelberg (2011)
2. IEEE standard for a precision clock synchronization protocol for networked measurement and control systems. IEEE Std 1588-2008 (Revision of IEEE Std 1588–2002), pp. 1–300 (2008)

3. Yang, Q., An, D., Yu, W.: On time desynchronization attack against IEEE 1588 protocol in power grid systems. In: Energytech, pp. 1–5 (2013)
4. Giorgi, G., Narduzzi, C.: Performance analysis of kalman-filter-based clock synchronization in IEEE 1588 networks. IEEE Trans. Instrum. Meas. **60**(8), 2902–2909 (2011)
5. Levesque, M., Tipper, D.: Improving the ptp synchronization accuracy under asymmetric delay conditions. In: IEEE International Symposium on Precision Clock Synchronization for Measurement, Control, and Communication, pp. 88–93 (2015)
6. Kovcshzy, T., Ferencz, B.: Performance evaluation of PTPd, a IEEE 1588 implementation, on the x86 linux platform for typical application scenarios. In: Instrumentation and Measurement Technology Conference, pp. 2548–2552 (2012)
7. Pravda, M., Lafata, P., Vodrazka, J.: Precise time protocol in ethernet over SDH network. In: International Conference on Telecommunications and Signal Processing, pp. 170–174 (2011)
8. Mahmood, A., Exel, R., Trsek, H., Sauter, T.: Clock synchronization over IEEE 802.11, a survey of methodologies and protocols. IEEE Trans. Ind. Inform. **PP**(99), 1–1 (2016)
9. Giorgi, G., Narduzzi, C.: Modeling and Simulation Analysis of PTP Clock Servo, pp. 155–161. In: IEEE International Symposium on Precision Clock Synchronization for Measurement, Control and Communication (2007)
10. Li, Q.: PTN clock synchronization technology and its applications. ZTE Commun. **08**(3), 27–31 (2010)
11. Correll, K., Barendt, N.: Design considerations for software only implementations of the IEEE 1588 precision time protocol. In: Conference on IEEE 1588 Standard for A Precision Clock Synchronization Protocol for Networked Measurement and Control Systems (2006)
12. Levesque, M., Tipper, D.: PTP++: A Precision Time Protocol Simulation Model for OMNeT++/INET (2015)
13. Eidson, J.C.: Measurement, control, and communication using IEEE 1588. Adv. Ind. Control. **111**(1), 75–82 (2006)
14. Bovet, D.P., Cesati, M.: Understanding the Linux Kernel. Oreilly Media (2000)

Influence Analysis of Combining Performance on Antenna Arraying upon the Residual Delay

Yiwen Jiao(✉), Yandu Liu, Hong Ma, and Tao Wu

Space Engineering University, Beijing 101416, China
avatar_jyw@163.com

Abstract. The theoretical method on the residual delay, in order to analyze the influence on antenna arraying signal combining, is given. First, we adopt a fractional delay all-pass filter to build a simulation model and give the design process of the simulation model on the residual delay. Then, we raise theoretical formula of the wideband signal combining loss on the residual delay through the derivation. Finally, we prove that theoretical results are consistent with the simulation results by simulation experiments and verify the correctness of the theoretical formula. Experimental results show that the greater the phase difference caused by the residual delay, the greater the impact of the wideband signal combining losses. For wideband signals, if the required combine loss is less than 0.1 dB, ε should be less than 0.068 symbol period.

Keywords: Antenna arraying · Wideband signal · Residual delay · Combining loss · Standard deviation

1 Introduction

With the deepening of deep space exploration, high precision image and video data transmission have become the inevitable requirement of the future deep space communication. The deep space TT&C system needs greater communication bandwidth and further communication distance. In recent years, NASA and ESA are planning the deep space TT&C network based on the new generation antenna array.

The loss of residual delay to wideband signal combine is one of the most important problems for large-scale broadband antenna arrays. For wideband signal combine, the influence of the residual delay that still exists in the array calibration will not be ignored, and the joint estimation of the residual time delay and phase must be considered. The residual delay will result in phase sliding at the edge of the bandwidth [1, 2]. For the same residual time delay, the greater the bandwidth, the greater the phase difference of the band edge, which will lose the coherence of the band edge, and then deteriorate the combine performance of the antenna array.

The quantitative analysis of the relationship between the residual time delay and the synthetic performance is of great significance for the study of related problems. The most widely used theoretical analysis method is the constraint condition method given by the document [1]. It is pointed out that the synthetic loss caused by residual delay can be ignored. The phase difference caused by residual delay is only a small part of the

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 694–703, 2019.
https://doi.org/10.1007/978-981-13-6264-4_82

whole period (for example, less than 5°). Although the constraint condition is simple and clear, it can only be used as a qualitative analysis. In document [3], the two antennas are taken as an example to analyze the random binary sequence from the time domain. It is pointed out that the loss of the residual delay caused by the residual delay is determined by the ratio of the residual delay to the symbol period. Although the relationship between the synthetic loss and the time delay is quantitatively analyzed, the factors considered are incomplete. For example, the influence of the shaping filter is not considered. In document [4], taking the two antennas for example, the synthetic signal is regarded as a multipath signal from the time domain angle. First, the lower bounds of the mean square error are determined by the residual time delay, and then the synthetic loss is calculated according to the relationship between the synthetic signal-to-noise ratio and the mean square error. The contribution of the literature is to give the theoretical lower bound of the influence of the residual delay on the synthetic performance, and point out that the synthetic performance of antenna array is affected by the roll drop coefficient of the SQRT-RC shaping pulse filter. The above literature shows that the larger the bandwidth is, the higher the accuracy of residual delay alignment is. The theoretical analysis of the literature [4] is more fully considered, but the shortcomings of above mentioned literature are obvious: (1) Only given the lower performance and the lack of quantitative theoretical deduction. (2) Only considering the delay difference between the two antennas, and the number of antennas in the actual array must be more, and the number of antennas must be considered. (3) Assuming that the time delay error is fixed, the residual time delay is random in the actual situation, and the analysis in the form of variance is more reasonable. In conclusion, the existing theoretical analysis methods for residual delay are still lack of accuracy and universality.

In this paper, an FSC system model considering residual delay is constructed, and a theoretical analysis method based on frequency domain integration is proposed. The method first analyses the influence of the residual time delay on the performance of the tone frequency signal, and then extends the integral to the influence on the performance of the wideband signal. Finally, the simulation is carried out.

2 FSC System Model Considering Residual Delay

2.1 System Model Construction

Figure 1 gives an FSC system model considering residual delay. For the separation problem, assuming that the carrier phase difference between signals is aligned, only the residual delay after delay compensation is considered.

As shown in Fig. 1, the source is a zero intermediate frequency baseband signal, and the amplitude is a constant 1, indicating the transmitter end's transmission filter, which has the square root rise cosine roll-down characteristic.

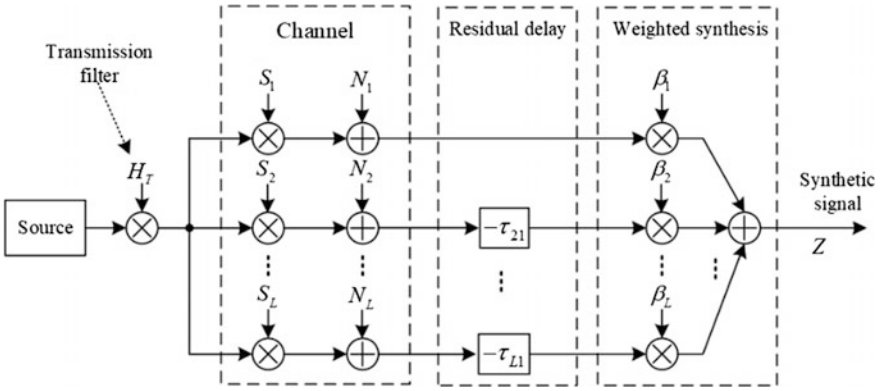


Fig. 1. FSC system model considering residual delay

$$|H_T(\omega)|^2 = \begin{cases} T_s & 0 \le |\omega| < (1-\alpha)\frac{\pi}{T_s} \\ \frac{T_s}{2}\left[1+\sin\frac{T_s}{2\alpha}\left(\frac{\pi}{T_s}-\omega\right)\right] & (1-\alpha)\frac{\pi}{T_s} \le |\omega| < (1+\alpha)\frac{\pi}{T_s} \\ 0 & |\omega| \ge (1+\alpha)\frac{\pi}{T_s} \end{cases} \tag{1}$$

where ω is angular frequency, T_s is a symbol period, and α is the coefficient of roll drop. The smaller the α is, the narrower the bandwidth is. But the ISI caused by sampling timing error is also more serious, so $0.2 \le \alpha \le 0.8$ is usually chosen. The number of received antennas is L, and the channel bandwidth is B, S_i represents the amplitude attenuation coefficient of the channel i channel, N_i represents additive Gauss white noise on the i channel, the bilateral power spectral density is N_{0i}. Considering the time delay between the corrected antenna and the reference antenna, the antenna No.1 is the reference antenna. As shown, $\tau_{11} = 0$ and the other antenna delay is $\tau_{i1}(i = 2, 3, \cdots, L)$. Considering the randomness of the residual delay and the independence of the antennas, according to the central limit theorem, $\tau_{i1}(i = 2, 3, \cdots, L)$ can be considered as an approximate Gauss distribution when the number of antennas is large. Therefore, $\tau_{i1}(i = 2, 3, \cdots, L)$ is assumed to obey the Gauss distribution with a mean value of 0 and a variance of $\sigma^2_{\tau i1}(i = 2, 3, \cdots, L)$. β_i represent the weighted value of the i path signal to maximize the signal-to-noise ratio (SNR), and the Z represents the output signal after weighted combine. In summary, synthetic signals based on this model can be expressed as

$$Z = \sum_{i=1}^{L} \beta_i [H_T S_i(t - \tau_{i1}) + N_i] \tag{2}$$

2.2 A Mathematical Model of Phase Difference Caused by Residual Delay

The residual delay will lead to phase sliding from low frequency to high frequency. The greater the bandwidth, the greater the phase difference between the band edges. Let ω_0

and ω_e denote the angular frequency of the center and edge of the baseband signal, respectively, and $\varphi_{i1}(\omega)$ indicates the phase difference of the angular frequency ω caused by the τ_{i1}. It is not difficult to see that even if the carrier phase $\varphi_{i1}(\omega_0) = 0$ exists, the phase of the other frequencies is not aligned due to the existence of the residual delay. And the greater the bandwidth, the greater the phase difference. The relationship between $\varphi_{i1}(\omega)$, ω and τ_{i1} can be expressed as follows:

$$\varphi_{i1}(\omega) = \begin{cases} 0 & i = 1 \\ \omega\tau_{i1} & i = 2, 3, \cdots, L \end{cases} \tag{3}$$

According to the properties of Gauss random variables, where $i = 2, 3, \cdots, L$, $\varphi_{i1}(\omega)$ obeys Gauss distribution. The mean value is 0, and the variance is

$$\sigma^2_{\varphi i1}(\omega) = \begin{cases} 0 & i = 1 \\ \sigma^2_{\tau i1}\omega^2 & i = 2, 3, \cdots, L \end{cases} \tag{4}$$

3 Theoretical Analysis of the Effect of Residual Delay on the Performance of Wideband Signal Combine

The baseband signal is a complex signal. According to the formula, the zero crossing bandwidth B can be expressed as

$$B = \frac{2\pi(1+\alpha)}{T_s} \tag{5}$$

According to the formula, the spectrum $R_i(\omega)$ of the useful signal received by the antenna i can be expressed as

$$R_i(\omega) = S_i H_T(\omega) \tag{6}$$

For $\beta_1 = 1$, to make the SNR of the synthesized signal Z the largest, $\beta_i (i = 2, \ldots, L)$ is calculated through the following formula.

$$\beta_i = \sqrt{\frac{R_i^2(\omega)}{R_1^2(\omega)}} \times \frac{N_{01}}{N_{0i}} = \frac{S_i N_{01}}{S_1 N_{0i}} \tag{7}$$

Because the noise N_i between the channels is independent of each other, the power of the noise component in the synthesized signal Z is

$$\delta_Z^2 = B \sum_{i=1}^{L} \beta_i^2 N_{0i} \tag{8}$$

According to Formula (2), the spectrum component of the synthesized signal Z is

$$Z_S(\omega) = \sum_{i=1}^{L} \beta_i S_i H_T(\omega) e^{j\varphi_{i1}(\omega)} \tag{9}$$

And power spectrum is

$$P_Z(\omega) = Z_S(\omega) Z_S^*(\omega) = \sum_{i=1}^{L} \sum_{j=1}^{L} \beta_i \beta_j S_i S_j |H_T(\omega)|^2 e^{j\Delta\varphi_{ij}(\omega)} \tag{10}$$

where $\Delta\varphi_{ij}(\omega) = \varphi_{i1}(\omega) - \varphi_{j1}(\omega)$, indicates the phase difference between the i and j antenna signals. There are four cases as follows:

$$\Delta\varphi_{ij}(\omega) = \begin{cases} 0 & i = j \\ \varphi_{i1}(\omega) & i \neq 1 \text{ and } j = 1 \\ -\varphi_{j1}(\omega) & j \neq 1 \text{ and } i = 1 \\ \varphi_{i1}(\omega) - \varphi_{j1}(\omega) & i \neq j \text{ and } i, j \neq 1 \end{cases} \tag{11}$$

Because $\Delta\varphi_{ij}(\omega)$ is also a Gauss random variable, so there are four cases of variance.

$$\sigma^2_{\Delta\varphi ij}(\omega) = \begin{cases} 0 & i = j \\ \sigma^2_{\tau i1}\omega^2 & i \neq 1 \text{ and } j = 1 \\ \sigma^2_{\tau j1}\omega^2 & j \neq 1 \text{ and } i = 1 \\ \left(\sigma^2_{\tau i1} + \sigma^2_{\tau j1}\right)\omega^2 & i \neq j \text{ and } i, j \neq 1 \end{cases} \tag{12}$$

So, the power of the useful signal component is

$$P(\omega) = E\left\{P_Z(\omega)\middle|\Delta\varphi_{ij}(\omega)\right\} = S_1^2 |H_T(\omega)|^2 \left(\sum_{i=1}^{L} \sum_{j=1}^{L} \gamma_i \gamma_j C_{ij}(\omega)\right) \tag{13}$$

where $\gamma_i \triangleq \frac{S_i^2}{S_1^2} \frac{N_{01}}{N_{0i}}$, $C_{ij}(\omega) \triangleq E\left[e^{j\Delta\varphi_{ij}(\omega)}\right]$.

According to the properties of Gauss random variable, when $i \neq j$, $C_{ij}(\omega)$ is actually the characteristic function of $\Delta\varphi_{ij}(\omega)$ [6]. Then

$$C_{ij}(\omega) = \begin{cases} 1 & i = j \\ e^{-\frac{\omega^2}{2}\sigma^2_{\tau i1}} & i \neq 1 \text{ and } j = 1 \\ e^{-\frac{\omega^2}{2}\sigma^2_{\tau j1}} & j \neq 1 \text{ and } i = 1 \\ e^{-\frac{\omega^2}{2}\left(\sigma^2_{\tau i1} + \sigma^2_{\tau j1}\right)} & i \neq j \text{ and } i, j \neq 1 \end{cases} \tag{14}$$

$C_{ij}(\omega)$ is the synthesized loss factor caused by the difference of the residual delay between the antenna i and antenna j. Under the ideal channel condition, $C_{ij}(\omega) \equiv 1, \forall i, j$, then

$$P_{ideal}(\omega) = S_1^2 |H_T(\omega)|^2 \left(\sum_{i=1}^{L} \sum_{j=1}^{L} \gamma_i \gamma_j \right) \tag{15}$$

The synthetic loss is defined as

$$D_{loss} \triangleq 10 \lg \frac{SNR_{real}}{SNR_{ideal}} \tag{16}$$

According Formulas (13) to (15), then

$$D_{loss}(\omega) = 10 \lg \left(\frac{\sum_{i=1}^{L} \gamma_i^2 + 2\gamma_1 \sum_{i=2}^{L} \gamma_i e^{-\frac{\omega^2}{2}\sigma_{\tau i1}^2} + \sum_{i=2}^{L} \sum_{\substack{j=2 \\ i \neq j}}^{L} \gamma_i \gamma_j e^{-\frac{\omega^2}{2}\left(\sigma_{\tau i1}^2 + \sigma_{\tau j1}^2\right)}}{\sum_{i=1}^{L} \sum_{j=1}^{L} \gamma_i \gamma_j} \right) \tag{17}$$

At this point, a theoretical formula for the effect of residual delay on the synthetic loss of tone frequency signals is given. For large-scale antenna arrays, it is generally considered that all antennas have the same performance and meet the conditions of uniform array. Then, the uniform array situation is considered. When $i = 2, 3, \cdots, L$, τ_{i1} is assigned to the Gauss distribution with a mean value of 0 and a variance of σ_τ^2. there is $S_i = S_1$, $N_{0i} = N_0$. For the uniform array, $\gamma_i = 1$, then

$$D_{loss}(\omega) = 10 \lg \left(\frac{1}{L} + \frac{2(L-1)}{L^2} e^{-\frac{\omega^2 \sigma_\tau^2}{2}} + \frac{(L-1)(L-2)}{L^2} e^{-\omega^2 \sigma_\tau^2} \right) \tag{18}$$

Integrating the synthesized signal power $P(\omega)$ and ideal signal power $P_{ideal}(\omega)$ in the bandwidth B, the power of wideband signal can be obtained. Based on this, the theoretical formula for the effect of residual delay on the synthetic loss of wideband signals can be obtained as follows.

$$D_{loss} = 10 \lg \left(\frac{\int_{B_L}^{B_H} |H_T(\omega)|^2 \left(\sum_{i=1}^{L} \gamma_i^2 + 2\gamma_1 \sum_{i=2}^{L} \gamma_i e^{-\frac{\omega^2}{2}\sigma_{\tau i1}^2} + \sum_{i=2}^{L} \sum_{\substack{j=2 \\ i \neq j}}^{L} \gamma_i \gamma_j e^{-\frac{\omega^2}{2}\left(\sigma_{\tau i1}^2 + \sigma_{\tau j1}^2\right)} \right) d\omega}{\int_{B_L}^{B_H} |H_T(\omega)|^2 \left(\sum_{i=1}^{L} \sum_{j=1}^{L} \gamma_i \gamma_j \right) d\omega} \right) \tag{19}$$

where B_H and B_L are the high-frequency and low-frequency boundaries of bandwidth B, respectively.

$$B_H = -B_L = \frac{B}{2} = \frac{(1+\alpha)\pi}{T_s} \tag{20}$$

A theoretical formula for the influence of uniform matrix residual delay on the synthetic loss of wideband signals is also given.

$$D_{loss} = 10\lg\left(\frac{\int_{B_L}^{B_H}|H_T(\omega)|^2\left(L+2(L-1)e^{-\frac{\sigma_\tau^2\omega^2}{2}}+(L-1)(L-2)e^{-\sigma_\tau^2\omega^2}\right)d\omega}{L^2\int_{B_L}^{B_H}|H_T(\omega)|^2 d\omega}\right) \tag{21}$$

According to Formula (1) and Formula (20), it can be obtained as

$$\int_{B_L}^{B_H}|H_T(\omega)|^2 d\omega = 2\pi \tag{22}$$

Then, Formula (21) can be written as

$$D_{loss} = 10\lg\left(\begin{array}{l}\frac{1}{L}+\frac{(L-1)}{\pi L^2}\int_{-\frac{(1+\alpha)\pi}{T_s}}^{\frac{(1+\alpha)\pi}{T_s}}|H_T(\omega)|^2 e^{-\frac{\sigma_\tau^2\omega^2}{2}}d\omega+\\ \frac{(L-1)(L-2)}{2\pi L^2}\int_{-\frac{(1+\alpha)\pi}{T_s}}^{\frac{(1+\alpha)\pi}{T_s}}|H_T(\omega)|^2 e^{-\sigma_\tau^2\omega^2}d\omega\end{array}\right) \tag{23}$$

Define $\varepsilon \triangleq \sigma_\tau/T_s$ as a code element error factor, and order $\theta = \omega T_s$ substituting Formula (23) for integral transformation.

$$D_{loss} = 10\lg\left(\begin{array}{l}\frac{1}{L}+\frac{(L-1)}{\pi L^2}\int_{-(1+\alpha)\pi}^{(1+\alpha)\pi}\left|H_T'(\theta)\right|^2 e^{-\frac{\varepsilon^2\theta^2}{2}}d\theta+\\ \frac{(L-1)(L-2)}{2\pi L^2}\int_{-(1+\alpha)\pi}^{(1+\alpha)\pi}\left|H_T'(\theta)\right|^2 e^{-\varepsilon^2\theta^2}d\theta\end{array}\right) \tag{24}$$

where

$$\left|H_T'(\theta)\right|^2 = \begin{cases}1 & 0\le|\theta|<(1-\alpha)\pi\\ \frac{1}{2}\left[1+\sin\left(\frac{\pi-\theta}{2\alpha}\right)\right] & (1-\alpha)\pi\le|\theta|<(1+\alpha)\pi\\ 0 & |\theta|\ge(1+\alpha)\pi\end{cases} \tag{25}$$

4 Simulation Results

First, the influence of residual delay on the synthetic loss of wideband signals is investigated under different rolling factor conditions. The simulation parameter $L = 100$, ε from 0 to 0.2, in the 0.01 step, α was 0.2, 0.5, and 0.8, respectively. And the order of the root liter cosine filter is 256. The simulation is divided into three steps

- Step1: The wideband signal is generated and divided into L paths, and the Gauss white noise is generated independently. The SNR is set to 10 dB.
- Step2: The L group fractional delay all-pass filter is designed by using the simulation model design method.
- Step3: The L signal is filtered through the L group filter and the L output signal is added together.
- Step4: The SNR estimation algorithm [7] is used to estimate the SNR of the synthesized signal, and the synthetic loss is calculated according to the formula (16).

The simulation results are compared with the theoretical values, as shown in Fig. 2.

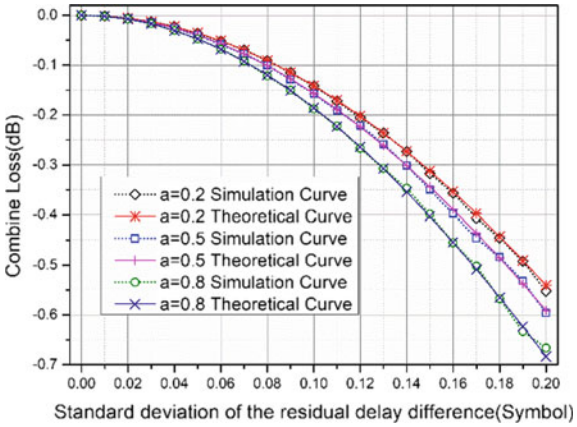


Fig. 2. Influence of residual delay on combine loss under different rolling factor

It can be seen from Fig. 2 that the simulation results are basically consistent with the theoretical results, and verify the correctness of the theoretical analysis. Theoretical and simulation results show that with the increase of residual delay, the loss of combine becomes larger and larger. In particular, when $\varepsilon \to \infty$ Formula (24) can be simplified to

$$\lim_{\varepsilon \to \infty} D_{loss}(L, \alpha, \varepsilon) = -10 \lg(L) \tag{26}$$

Next, the influence of residual delay on the synthetic loss of wideband signals is simulated under different antenna numbers. Figure 3 gives the simulation and theoretical curves of the variation of D_{loss} with ε when $\alpha = 0.5$, L are 10, 25 and 100, respectively.

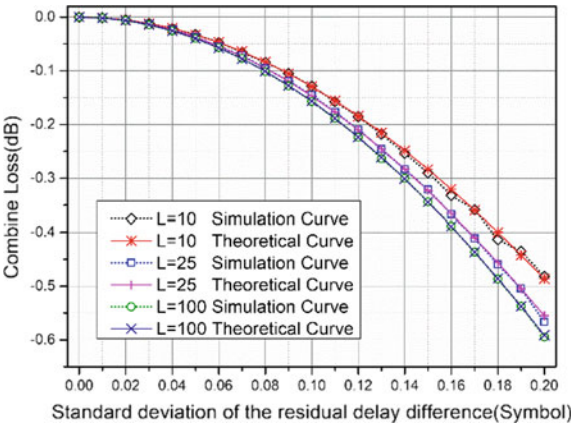


Fig. 3. Influence of residual delay on combine loss under different antenna numbers

As shown in Fig. 3, the simulation results are basically consistent with the theoretical results, which verify the correctness of the theoretical analysis. The greater the L, the greater the loss of combine. When $\alpha = 1$ and $L \to \infty$, Formula (24) can be simplified to

$$\lim_{\substack{\alpha = 1 \\ L \to \infty}} D_{loss}(L, \alpha, \varepsilon) = 10 \lg \left(\frac{1}{4\sqrt{\pi}} \frac{erf(2\pi\varepsilon)}{\varepsilon} + \frac{1}{4\pi} \int_{-2\pi}^{2\pi} \cos\left(\frac{\theta}{2}\right) e^{-\varepsilon^2 \theta^2} d\theta \right) \tag{27}$$

where, $erf(\bullet)$ is an error function. In practical applications, the upper limit ε determined by the Formula (27). For example, for wideband signals, if the required combine loss is less than 0.1 dB, ε should be less than 0.068 symbol period. According to Formula (20), if the bandwidth is 1 GHz, $\alpha = 1$, then the symbol period $T_s = 2$ ns. From $\varepsilon \triangleq \sigma_\tau / T_s$, the standard deviation of the residual time delay distribution should be less than 0.136 ns.

5 Conclusion

In this paper, the influence of the residual delay on the combined performance of the broadband signal of the antenna array is quantitatively analyzed by establishing a model, mathematical derivation, and simulation verification. The system block diagram of the full spectrum synthesis model with residual time delay is given, and the residual

time delay distribution model is constructed. The theoretical formula of the effect of residual time delay on the synthesis loss of broadband signals under uniform array is deduced. The theoretical formula has been obtained and the deficiency of theoretical research has been perfected.

References

1. Rogstad, D.H., Mileant, A., Pham, T.T.: Antenna Arraying Techniques in the Deep Space Network. John Wiley & Sons Inc., Hoboken, New Jersey (2003)
2. Shen, C.Y.: Research on the Combining Techniques on Joint Reception for Multi-antenna Signals. PLA Information Engineering University for the Degree of Doctor of Engineering, Zhengzhou (2011). (in Chinese)
3. Jiao, Q.J., Liu, Y.Y., Chen J.: Research on the compensation method of time delay upon antenna arraying wideband signal. In: The 26th Conference of Chinese spacecraft TT&C technology in Chinese Professional Committee of Astronautics Aircraft, pp. 365–368 (2012). (in Chinese)
4. Shen, C.Y.: Research on the Combining Techniques on Joint Reception for Multi-antenna Signals. PLA Information Engineering University for the Degree of Doctor of Engineering, Zhengzhou (2011). (in Chinese)
5. Thiran, J.: Recursive digital filters with maximally flat group delay. IEEE Trans. Circuit Theor. **18**(6), 659–663 (1971)
6. Zhao, S.J., Zhao, J.X.: Signal Detection and Estimation Theory. Tsinghua University Press, Beijing (2005). (in Chinese)
7. Cesarone, R.J., Abraham, D.S., Deutsch, L.J.: Prospects for a next-generation deep-space network. Proc. IEEE **95**(10), 1902–1915 (2007)
8. Torre, A.: Wideband Antenna Arraying Over Long Distances. California: JPL IPN Progress Report 42–194 (2013)
9. Marti, S.: Antenna arrays for high data return from future deep space missions. In: Stockholm, Sweden: 2012 The 12th International Conference on Space Operations (2012)
10. Bagri, D.S., Statman, J.I., Gatti, M.S.: Proposed array-based deep space network for NASA. Proc. IEEE **95**(10), 1916–1922 (2007)
11. Soriano, M., Rogstad, S., Navarro, R., et al.: Design and implementation of a deep space communications complex downlink array. In: Big Sky, MT, United States: 2014 IEEE Aerospace Conference (2014)

Digital Precoding Design for MIMO System with One-Bit ADC Receivers

Qitong Hou, Rui Wang(✉), Erwu Liu, and Dongliang Yan

School of Electronics and Information Engineering, Tongji University, Shanghai, China
{houqitong,ruiwang,yandl}@tongji.edu.cn, erwu.liu@ieee.org

Abstract. When wireless communication transmits in the bandwidth of gigahertz, radio-frequency (RF) chains and high-resolution analog-to-digital converters (ADCs) increasingly bring burdens to energy consumption. To solve this problem, promising techniques are to use low-resolution ADCs and hybrid precoding architectures. In this work, we combine these solutions for reducing the power cost more efficiently in the multi-input and multi-output (MIMO) system. The aim is designing the optimal digital precoder, whose objective is to increase the achievable rate. For this purpose, the expression for the achievable rate according to Bussgang theorem in flat-fading channels was first derived. To obtain the optimal digital precoder, we solve the optimal digital precoding matrix form using singular value decomposition (SVD) method. We then transfer the digital precoding matrix design problem into optimization problem according to power allocation, whose optimal closed-form solution can be found with the usage of Karush–Kuhn–Tucker (KKT) conditions. Our simulation results reveal that this proposed algorithm increases the achievable rates of the system effectively.

Keywords: Multi-input and multi-output system · One-bit analog-to-digital converter · Hybrid precoding architecture

1 Introduction

Multi-input and multi-output (MIMO) system is an important part in the fifth-generation networks, which is an effective way to improve the transmission reliability and spectral efficiency [1]. However, the increasing number of antennas brings difficulties for power consumption and hardware costing, making the solutions for high-resolution ADCs and fully digital precoding architectures difficult to implement in practice [2]. To address this problem, using one-bit ADCs [3,4] and hybrid precoding architectures [5] are two promising solutions. Nevertheless, the two solutions represent two extremes that is to decrease either RF chains or ADC bits, and they are not able to effectively save the power consumption or hardware costing. The aim is to investigate the combining architecture of these two solutions.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 704–712, 2019.
https://doi.org/10.1007/978-981-13-6264-4_83

This paper investigates the digital precoder design problem for MIMO system, whose receiver uses one-bit ADCs. The work derives the achievable rate in the scenario of flat-fading channels, which on the basis of Bussgang theorem [6]. Moreover, in order to get the optimal digital precoder matrix, we are able to obtain the optimal matrix form using singular value decomposition (SVD) method when we fix the analog precoder matrix. Moreover, we transfer the digital precoder matrix design optimization problem into an optimization problem according to power allocation, and we can solve the power allocation problem with the usage of Karush–Kuhn–Tucker (KKT) conditions. Simulation reveals that this proposed algorithm can increase the capacity of the channel significantly.

2 System Model

We propose the system model for the MIMO system with one-bit ADCs in this section, which is shown in Fig. 1. N_t and N_r antennas are, respectively, equipped at the transmitter and receiver. N_{RF}^t RF chains are assumed at the transmitter, satisfying $N_{RF}^t \leq N_t$. Transmit and receive antennas collect and send N_s data streams, which satisfy $N_s \leq N_{RF}^t$.

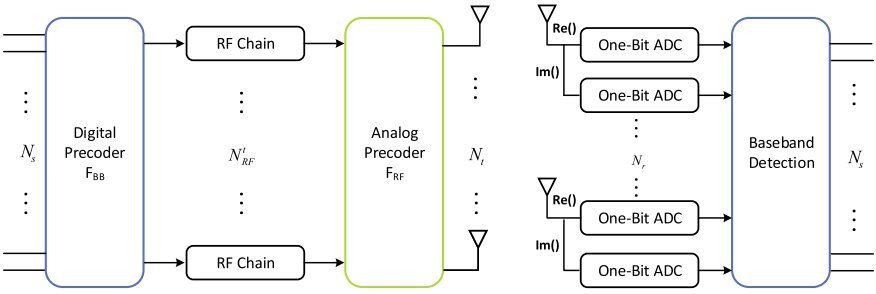


Fig. 1. System model

Assuming the scenario of flat-fading channels, the receive signal is

$$\mathbf{y} = \mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s} + \mathbf{n}. \tag{1}$$

$\mathbf{y} \in \mathbb{C}^{N_s \times 1}$ is the receive signal before the quantization process, $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$ satisfies $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \frac{P_t}{N_s}\mathbf{I}_{N_s}$, and the transmission power is P_t, the white Gaussian noise satisfies $\mathbf{n} \sim \mathcal{CN}(0, \mathbf{I}_{N_r})$, the flat-fading channel is $\mathbf{H} \in \mathbb{C}^{N_s \times N_t}$, $\mathbf{F}_{RF} \in \mathbb{C}^{N_t \times N_{RF}^t}$ and $\mathbf{F}_{BB} \in \mathbb{C}^{N_{RF}^t \times N_s}$ are the analog precoding matrix and digital precoding matrix, satisfying the power constraint $\|\mathbf{F}_{RF}\mathbf{F}_{BB}\|_F^2 \leq P_t$.

After the quantization process, the signal is

$$\mathbf{r} = \mathcal{Q}(\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s} + \mathbf{n}). \tag{2}$$

$\mathcal{Q}(\cdot)$ is denoted as the quantization function quantizing the real and imaginary parts separately, which can be expressed by $\mathcal{Q} = \frac{1}{\sqrt{2}} sign(\Re(\cdot)) + \frac{1}{\sqrt{2}} j sign(\Im(\cdot))$. The quantization process outputs in $\mathcal{R} = \{\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}j, \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}j, -\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}j, -\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}j\}$.

3 Achievable Rate

3.1 Bussgang Theorem

The quantization process for one-bit ADCs is nonlinear, but we are able to reformulate the process as an equivalent linear one with the usage of Bussgang theorem. In particular, the Bussgang theorem for (2) is

$$\mathbf{r} = \mathbf{AHF}_{RF}\mathbf{F}_{BB}\mathbf{s} + \widetilde{\mathbf{n}}. \tag{3}$$

$\mathbf{A}$ is the linear coefficient. The equivalent quantized noise is $\mathbf{q}$. The total noise is $\widetilde{\mathbf{n}} = \mathbf{An} + \mathbf{q}$, which is comprised of quantization noise and Gaussian noise.

One significant choice for $\mathbf{A}$ is minimizing the power of $\mathbf{q}$, which yields that $\mathbf{q}$ does not correlate with $\mathbf{y}$. Thus, $\mathbf{A}$ can be obtained by

$$\arg\min_{\mathbf{A}} \mathbb{E}\{\|\mathbf{q}\|_2^2\} = \arg\min_{\mathbf{A}} \mathbb{E}\{\|\mathbf{r} - \mathbf{Ay}\|_2^2\}, \tag{4}$$

whose solution is given by

$$\mathbf{A} = \mathbf{R}_{\mathbf{yr}}^H \mathbf{R}_{\mathbf{y}}^{-1}, \tag{5}$$

where $\mathbf{R}_{\mathbf{yr}}$ is the cross-correlation of $\mathbf{y}$ and $\mathbf{r}$, and $\mathbf{R}_{\mathbf{y}}$ is the autocorrelation of $\mathbf{y}$. According to Gaussian input signals and one-bit ADCs, $\mathbf{R}_{\mathbf{yr}}$ is

$$\mathbf{R}_{\mathbf{yr}} = \sqrt{\frac{2}{\pi}} \mathbf{R}_{\mathbf{y}} diag(\mathbf{R}_{\mathbf{y}})^{-\frac{1}{2}}. \tag{6}$$

Substituting (6) into (5), we can get $\mathbf{A}$ is only related to $\mathbf{R}_{\mathbf{y}}$:

$$\mathbf{A} = \sqrt{\frac{2}{\pi}} diag(\mathbf{R}_{\mathbf{y}})^{-\frac{1}{2}}. \tag{7}$$

In reality, we can reconstruct $\mathbf{R}_{\mathbf{y}}$ with the usage of the same method provided in [7]. Thus, we can approximate the matrix $\mathbf{A}$ as

$$\mathbf{A} \cong \sqrt{\frac{2}{\pi}} \sqrt{\frac{1}{1+P_t}} \mathbf{I}_{N_r} = \alpha \mathbf{I}_{N_r}, \tag{8}$$

where $\alpha = \sqrt{\frac{2}{\pi}} \sqrt{\frac{1}{1+P_t}}$.

3.2 Achievable Rate for the System

To get the achievable rate, which correlated with $\mathbf{R}_{\tilde{\mathbf{n}}}$, we need to derive $\mathbf{R}_{\mathbf{q}}$ relating to $\mathbf{R}_{\mathbf{r}}$.

As shown in [8], we can use the arcsine law to obtain $\mathbf{R}_{\mathbf{r}}$:

$$\begin{aligned}\mathbf{R}_{\mathbf{r}} = \frac{2}{\pi}(&arcsin(diag(\mathbf{R}_{\mathbf{y}})^{-\frac{1}{2}}\Re(\mathbf{R}_{\mathbf{y}})diag(\mathbf{R}_{\mathbf{y}})^{-\frac{1}{2}}) \\ &+ jarcsin(diag(\mathbf{R}_{\mathbf{y}})^{-\frac{1}{2}}\Im(\mathbf{R}_{\mathbf{y}})diag(\mathbf{R}_{\mathbf{y}})^{-\frac{1}{2}})).\end{aligned} \tag{9}$$

While the quantization noise in most cases cannot be Gaussian distributed, we can model the quantization noise which is distributed in Gaussian for the achievable rates, since the worst case is Gaussian. In particular, we can use the model of $\mathbf{q}$ being white Gaussian noise to get the achievable rate:

$$\begin{aligned}\mathbf{R}_{\mathbf{q}} &= \mathbf{R}_{\mathbf{r}} - \mathbf{A}\mathbf{R}_{\mathbf{y}}\mathbf{A}^H \\ &= \frac{2}{\pi}(arcsin(\mathbf{M}) + jarcsin(\mathbf{N})) - \frac{2}{\pi}(\mathbf{M} + j\mathbf{N}),\end{aligned} \tag{10}$$

in which we define

$$\mathbf{M} = diag(\mathbf{R}_{\mathbf{y}})^{-\frac{1}{2}}\Re(\mathbf{R}_{\mathbf{y}})diag(\mathbf{R}_{\mathbf{y}})^{-\frac{1}{2}}, \tag{11}$$

$$\mathbf{N} = diag(\mathbf{R}_{\mathbf{y}})^{-\frac{1}{2}}\Im(\mathbf{R}_{\mathbf{y}})diag(\mathbf{R}_{\mathbf{y}})^{-\frac{1}{2}}. \tag{12}$$

We can apply the arcsine law and approximate (10) as

$$\mathbf{R}_{\mathbf{q}} \cong (1 - \frac{2}{\pi})\mathbf{I}_{N_r}, \tag{13}$$

as a result the quantized noise can be approximated as uncorrelated whose variance is $1 - \frac{2}{\pi}$.

According to a Gaussian input and the Bussgang theorem, we can obtain

$$\mathbf{R}_{\mathbf{y}} = \frac{P_t}{N_s}\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{F}_{BB}^H\mathbf{F}_{RF}^H\mathbf{H}^H + \mathbf{I}_{N_r}. \tag{14}$$

Moreover, using $\mathbf{A}$ as in (8), $\mathbf{q}$ is uncorrelated with $\mathbf{n}$. Thus, we can get $\mathbf{R}_{\tilde{\mathbf{n}}}$ by

$$\begin{aligned}\mathbf{R}_{\tilde{n}} &= \mathbf{A}\mathbf{A}^H + \mathbf{R}_{\mathbf{q}} \\ &\cong (\alpha^2 + 1 - \frac{2}{\pi})\mathbf{I}_{N_r}.\end{aligned} \tag{15}$$

Thus the achievable rate is

$$C = log_2|\mathbf{I}_{N_s} + \frac{P_t}{N_s}\mathbf{R}_{\tilde{\mathbf{n}}}^{-1}\mathbf{A}\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{F}_{BB}^H\mathbf{F}_{RF}^H\mathbf{H}^H\mathbf{A}^H|. \tag{16}$$

After substituting (8), (14) and (15) into (16), the achievable rate is

$$C = log_2|\mathbf{I}_{N_s} + \frac{P_t}{N_s}\frac{\pi\alpha^2}{\pi\alpha^2 + \pi - 2}\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{F}_{BB}^H\mathbf{F}_{RF}^H\mathbf{H}^H|. \tag{17}$$

4 Problem Formulation

In this section, we obtain the digital precoding design problem with the objective of increasing the achievable rate by designing the digital baseband precoder $\mathbf{F}_{BB}$ with $\mathbf{F}_{RF}$ fixed.

The optimization design problem for $\mathbf{F}_{BB}$ is formulated as

$$\begin{aligned} \max_{\mathbf{F}_{BB}} \quad & C \\ \text{s.t.} \quad & trace\{\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{F}_{BB}^{H}\mathbf{F}_{RF}^{H}\} \leq P_t, \end{aligned} \tag{18}$$

where we define $\mathbf{X} = \mathbf{F}_{RF}^{H}\mathbf{H}^{H}\mathbf{H}\mathbf{F}_{RF}$.

Lemma 1. *The optimal digital precoder structure is* $\mathbf{F}_{BB} = \mathbf{X}^{-\frac{1}{2}}\mathbf{U}\mathbf{\Sigma}^{\frac{1}{2}}$, *where* $\mathbf{\Sigma}$ *is diagonal and* $\mathbf{U}$ *is para-unitary from the singular value decomposition of* $\mathbf{X}^{-\frac{1}{2}}\mathbf{F}_{RF}^{H}\mathbf{F}_{RF}\mathbf{X}^{-\frac{1}{2}}$.

Proof. For (18), the optimal solution for digital precoder makes $\mathbf{F}_{BB}^{H}\mathbf{X}\mathbf{F}_{BB}$ diagonal. Thus, by assuming $\mathbf{F}_{BB}^{H}\mathbf{X}\mathbf{F}_{BB} = \mathbf{\Sigma}$, in which $\mathbf{\Sigma}$ is diagonal with the elements increasing order arranged, we can get $\mathbf{X}^{\frac{1}{2}}\mathbf{F}_{BB} = \mathbf{U}\mathbf{\Sigma}^{\frac{1}{2}}$, where $\mathbf{U} \in \mathbb{C}^{N_{RF}^{t} \times N_s}$ is para-unitary. And we can get the optimal form of $\mathbf{F}_{BB}$ as

$$\mathbf{F}_{BB} = \mathbf{X}^{-\frac{1}{2}}\mathbf{U}\mathbf{\Sigma}^{\frac{1}{2}}. \tag{19}$$

Substituting the optimal $\mathbf{F}_{BB}$ structure (19) into (18), we can get

$$\begin{aligned} & trace\{\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{F}_{BB}^{H}\mathbf{F}_{RF}^{H}\} \\ =& trace\{\mathbf{F}_{BB}^{H}\mathbf{F}_{RF}^{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\} \\ =& trace\{\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{U}^{H}\mathbf{X}^{-\frac{1}{2}}\mathbf{F}_{RF}^{H}\mathbf{F}_{RF}\mathbf{X}^{-\frac{1}{2}}\mathbf{U}\mathbf{\Sigma}^{\frac{1}{2}}\} \\ \geq& trace\{\mathbf{\Sigma}\tilde{\mathbf{\Sigma}}'\}, \end{aligned} \tag{20}$$

where $\mathbf{X}^{-\frac{1}{2}}\mathbf{F}_{RF}^{H}\mathbf{F}_{RF}\mathbf{X}^{-\frac{1}{2}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{U}}^{H}$, and the singular values in $\tilde{\mathbf{\Sigma}}$ are arranged in increasing order, then we select the N_s largest singular values and order them in decreasing as the elements of $\tilde{\mathbf{\Sigma}}' \in \mathbb{C}^{N_s \times N_s}$ in diagonal. And $\tilde{\mathbf{\Sigma}}'$ is defined by

$$\tilde{\mathbf{\Sigma}}' = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_{N_s} \end{bmatrix} \tag{21}$$

Note that in (20) the equality holds when $\mathbf{U} = \tilde{\mathbf{V}}'$, and $\tilde{\mathbf{V}}'$ is the singular vector matrix, which corresponds to the singular values of $\tilde{\mathbf{\Sigma}}'$.

As a result, the optimal digital precoder is $\mathbf{F}_{BB} = \mathbf{X}^{-\frac{1}{2}}\tilde{\mathbf{V}}'\mathbf{\Sigma}^{\frac{1}{2}}$.

Another optimization problem can be formulated to solve $\mathbf{\Sigma}$. First, $\mathbf{\Sigma}$ can be defined by

$$\mathbf{\Sigma} = \begin{bmatrix} \lambda_{q,1} & & & \\ & \lambda_{q,2} & & \\ & & \ddots & \\ & & & \lambda_{q,N_s} \end{bmatrix} \tag{22}$$

Then for $\lambda_{q,i}, \forall i$, the optimization problem can be written as

$$\begin{aligned} \min_{\lambda_{q,i} \geq 0} \quad & -\sum_{i=1}^{N_s} log_2(1+\gamma\lambda_{q,i}) \\ \text{subject to} \quad & \sum_{i=1}^{N_s} \lambda_{q,i}\lambda_i \leq P_t. \end{aligned} \tag{23}$$

Lemma 2. *The optimization design problem* (23) *is convex.* $\lambda_{q,i}$ *is denoted by the ith diagonal element, thus the optimal* $\lambda_{q,i}$ *is* $\lambda_{q,i} = [\frac{1}{\lambda_i \beta \ln 2} - \frac{1}{\gamma}]^{\dagger}$, *where* β *is the Lagrangian multiplier and* $\gamma = \frac{\pi\alpha^2}{\pi\alpha^2+\pi-2}$.

Proof. For (23), the objective function can be seen as $k = m(n(\lambda_{q,i}))$, where $m(x) = log_2(x)$ and $n(\lambda_{q,i}) = 1+\gamma\lambda_{q,i}$. For the reason that $m(x)$ is nondecreasing concave and $n(\lambda_{q,i})$ is concave according to $\lambda_{q,i}$, thus $m(n(\lambda_{q,i}))$ is concave [10].

Since (23) is a convex problem, we can use KKT conditions to find the optimal solution. First, for (23), we formulate the Lagrangian function:

$$\mathcal{L} = -\sum_{i=1}^{N_s} log_2(1+\gamma\lambda_{q,i}) + \beta(\sum_{i=1}^{N_s} \lambda_{q,i}\lambda_i - P_t) - \sum_{i=1}^{N_s} \beta_i\lambda_{q,i}, \tag{24}$$

in which Lagrangian multipliers are β and β_i. And the solution set for KKT conditions is obtained by

$$\frac{\partial \mathcal{L}}{\partial \lambda_{q,i}} = -\frac{\gamma}{1+\gamma\lambda_{q,i}} \cdot \frac{1}{\ln 2} + \beta\lambda_i - \beta_i = 0 \tag{25a}$$

$$\beta(\sum_{i=1}^{N_s} \lambda_{q,i}\lambda_i - P_t) = 0 \tag{25b}$$

$$\beta_i\lambda_{q,i} = 0 \tag{25c}$$

$$\beta \geq 0 \tag{25d}$$

$$\beta_i \geq 0, \forall i. \tag{25e}$$

By multiplying (25a) by $\lambda_{q,i}$, we can get

$$\lambda_{q,i}(\lambda_i\beta - \frac{1}{\ln 2}\frac{\gamma}{1+\gamma\lambda_{q,i}}) = \lambda_{q,i}\beta_i = 0. \tag{26}$$

We can discuss two cases satisfying (26):
Case 1: When $\beta > \frac{\gamma}{\lambda_i \ln 2}$, we can get $\lambda_{q,i} = 0$.

Case 2: Otherwise, when $\beta \leq \frac{\gamma}{\lambda_i \ln 2}$, since $\beta_i \geq 0$, it can be obtained $\lambda_{q,i} > 0$ and $\beta = \frac{1}{\lambda_i \ln 2} \frac{\gamma}{1+\gamma\lambda_{q,i}}$. For this case, $\lambda_{q,i}$ optimal solution is $\lambda_{q,i} = \frac{1}{\lambda_i \beta \ln 2} - \frac{1}{\gamma}$.

We combine the result for the two cases and obtain $\lambda_{q,i}$ solution, which is

$$\lambda_{q,i} = [\frac{1}{\lambda_i \beta \ln 2} - \frac{1}{\gamma}]^{\dagger}. \tag{27}$$

We can use bisection method to solve the upper bound of β. The inequality in (25d) requires that β is larger than zero, thus we can select zero to be the lower bound of β. Furthermore, the equality of $\sum_{i=1}^{N_s} \lambda_{q,i}\lambda_i = P_t$ should be satisfied by β. We can get the constraint of β for λ_i

$$\sum_{i=1}^{Ns} [\frac{1}{\beta \ln 2} - \frac{\lambda_i}{\gamma}]^{\dagger} = P_t. \tag{28}$$

From the inequality of $\sum_{i=1}^{Ns} \frac{1}{\beta \ln 2} - \frac{\lambda_{N_s}}{\gamma} \geq P_t$, we can get the upper bound of β. The upper bound for β is

$$\beta_{max} = \frac{1}{\ln 2} \frac{\gamma N_s}{\gamma P_t^2 + N_s \lambda_{N_s}}. \tag{29}$$

5 Simulation Results

The performance for the proposed algorithm in different SNRs at $N_t = 6$, $N_r = 4$, and $N_s = 2$ was investigated in this section. From Fig. 2, it can be known that the

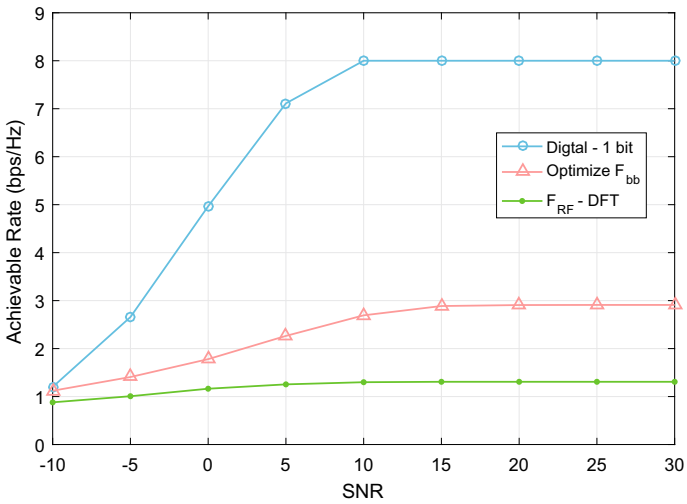


Fig. 2. Achievable Rates versus SNR

algorithm discussed can efficiently improve the achievable rates of the system in all SNRs. Further, the quantization process impedes increasing for the achievable rate with SNR increasing, as a result the achievable rate will get saturated finally.

Moreover, we compare with the fully digital precoding architecture, which uses one-bit ADCs in the receiver. With $N_r = 4$, there exists no more than $2^{2N_r} = 16$ possible quantization outputs, thus the upper bound of the achievable rates is $2N_r = 8$ bps/Hz.

6 Conclusion

In this paper, digital precoding design optimization problem for MIMO system with one-bit ADCs is investigated. An expression for the achievable rates is derived with the usage of the Bussgang Theorem, and we formulate the optimization problem for obtaining the maximum of the achievable rates. To optimize the digital precoder matrix, the optimal form of the digital precoder is obtained with the usage of SVD method. The digital precoding design optimization problem is transferred into an optimization problem according to power allocation, which is able to be solved by KKT conditions. Moreover, we use the bisection method for determining the parameters. Finally, our simulation results have clearly shown that the proposed method is effective.

Acknowledgement. This paper was supported by the National Science Foundation China under Grant 61771345.

References

1. Wang, R., Yuan, X., Yeung, R.W.: Mimo multipair two-way relaying with distributed relays: joint signal alignment and interference neutralization. IEEE Trans. Inf. Theory **62**(3), 1326–1343 (2016)
2. Murmann, B.: Adc performance survey 1997-2010 (2015). http://www.stanford.edu/murmann/adcsurvey.html
3. Mo, J., Heath, R.W.: Capacity analysis of one-bit quantized mimo systems with transmitter channel state information. IEEE Trans. Signal Process. **63**(20), 5498–5512 (2015)
4. Li, Y., Tao, C., Liu, L., Seco-Granados, G., Swindlehurst, A.L.: Channel estimation and uplink achievable rates in one-bit massive mimo systems. In: 2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), pp. 1–5. IEEE (2016)
5. Yu, X., Shen, J.-C., Zhang, J., Letaief, K.B.: Alternating minimization algorithms for hybrid precoding in millimeter wave mimo systems. IEEE J. Sel. Top. Signal Process. **10**(3), 485–500 (2016)
6. Bussgang, J.J.: Crosscorrelation Functions of Amplitude-Distorted Gaussian Signals (1952)
7. Bar-Shalom, O., Weiss, A.J.: DOA estimation using one-bit quantized measurements. IEEE Trans. Aerosp. Electron. Syst. **38**(3), 868–884 (2002)

8. Jacovitti, G., Neri, A.: Estimation of the autocorrelation function of complex gaussian stationary processes by amplitude clipped signals. IEEE Trans. Inf. Theory **40**(1), 239–245 (1994)
9. Wang, R., Tao, M., Xiang, Z.: Nonlinear precoding design for mimo amplify-and-forward two-way relay systems. IEEE Trans. Veh. Technol. **61**(9), 3984–3995 (2012)
10. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)

Mobile Communications

A Social-Aware Mode Selection Algorithm for D2D Communications

Wei-feng Lu[1,3(✉)], Wei-jun Lin[1,3], Si-guang Chen[2], and Li-jun Yang[2]

[1] School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
luwf@njupt.edu.cn
[2] College of IoT, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
[3] Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, Nanjing 210023, China

Abstract. Aiming at mode selection in D2D communication, a mode selection strategy based on social-aware was proposed. In addition to physical connection status, users' social relationships also served as an observation element. Through relationship between users, the closeness of user relationship can be judged, and then through users' reference, the similarity between users can be obtained. In this way, by combining physical connection with social relationships, users' aggregate utility was acquired. According to the physical connection quality and the social relationship strength, taking maximizing users' aggregate utility as the goal, a random algorithm was designed to match users. Simulation results show that mode selection based on social perception can effectively improve users' aggregate utility.

Keywords: D2D communication · Mode selection · Social aware

1 Introduction

With the popularization of Internet technology and smartphone, Internet plays a more and more important role in people's daily life, network transmission bears higher and higher load and current communication technology cannot meet people's demands on network communication [1]. Currently, more and more enterprises and scientific research institutions, such as Qualcomm, Ericsson, and Bell Laboratory, have started the research of fifth-generation mobile communication technology. As the important part of 5G technology, D2D (Device-to-Device) communication technology allows users to transmit data without forwarding through base station [2], which not only effectively improves the use ratio of spectrum resources but lowers network delay. D2D communication can be divided by spectrum resources into out-of-band D2D mode and in-band D2D mode. The D2D users using in-band mode can communicate with each other through the authorized frequency band of cellular network and the D2D users using out-of-band mode can communicate with each other through unauthorized frequency band. In-band D2D can be divided into the multiplexed mode and dedicated mode. The D2D users using multiplexed mode share spectrum resources with cellular

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 715–724, 2019.
https://doi.org/10.1007/978-981-13-6264-4_84

users leading to their mutual intervention; the D2D users using dedicated mode communicate with each other through the mutually orthogonal spectrum resources without suffering from the intervention of cellular users.

The traditional D2D communication mode generally considers the quality of the physical link in communication mode selection strategy. Literature [3] raises a distance-based D2D mode selection strategy and determines whether to use multiplexed mode in D2D communication according to the distance between D2D users and the cellular users of base station and multiplexed resources. The plan is helpful for reducing the mutual intervention of D2D users and cellular users. The algorithm has simple investigation factors and adopts the strategy of random selection in choosing the objects for multiplexed resources, which greatly lowers optimization effect. D2D communication is after all the communication among humans, so it is necessary to regard the social property of humans as the research factor [4]. Social-aware-based D2D mode selection strategy inspects the social-aware among users and determines whether to build D2D connection by quantifying the relational closeness among users [5–7]. Literature [8] raises an optimum community perception resources allocation algorithm, which takes advantage of the social network property of community to reduce the transmission time in D2D communication and achieves the allocation of resources by considering the social features of users. The centralized control of the plan adds to the loads of base station. Literature [9] designs a social-aware-based D2D video broadcast system which adopts a cooperative game-based distributed algorithm which effectively improves the performance of the network but has a high complexity. Literature [10] transforms mode selection and resources allocation into a non-transferable coalition game, raises a distributed coalition formation algorithm and takes advantage of the algorithm to obtain the joint strategy of mode selection and link allocation and prove the stability of coalition structure. Literature regards user efficiency as optimization measurement, but power consumption is set as a constant value. In addition, the influences of social relation only works as the model of negative exponent function. In contrast, this paper raises a social-aware-based D2D mode selection algorithm, which measures the social relation of users according to the cosine similarity and relational closeness among users, obtains the comprehensive income of users by combining the quality of physical link and regards comprehensive income as an optimization goal in mode selection. The simulation results show that the algorithm can effectively improve the comprehensive income of users.

The remainder of this paper is organized as follows. In Sect. 2, the system model is proposed. In Sect. 3, a social-aware mode selection algorithm for D2D Communications is analyzed. Simulation results are illuminated in Sect. 4 and conclusions are given in Sect. 5.

2 System Model

The research is based on the OFDMA cellular system model, shown in Fig. 1. The system model consists of a base station, M content senders and N content receivers. User's equipment has many groups of reception and transmission antenna and district information channel obeys single path loss. District users have two communication

modes to choose, namely cellular mode and D2D mode. When content reception users succeed in D2D match, dedicated resources are adopted in D2D communication. When D2D match fails, cellular mode is used to obtain content from the base station. In D2D mode, one receiver only can obtain data from one sender and one sender can transmit data to multiple receivers.

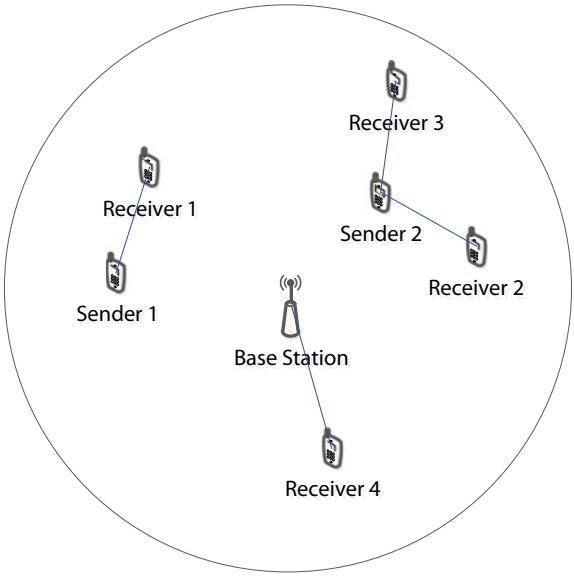


Fig. 1. System model

3 Mode Selection

There are much hot spot information in life and many users ask for them. If they obtain these information from base station each time it will cause the wastage of network resources. If users ask the users who have obtained content for information, the use ratio of internet resources can be improved obviously. However, users don't want to provide content for other users because that will consume their calculation resources and electricity of equipment, and they also need other users to provide content for themselves. The paper achieves the simulation of users from two aspects. First, users tend to provide services for their friends. Introduce relational closeness as a measurement and determine the relation among users by calculating the number of friends. Assuming fi is the collection of friends of user i, the relational closeness $F_{i,j}$ between users i and j can be expressed as

$$F_{i,j} = \frac{|f_i \cap f_j|}{|f_i \cup f_j|} \tag{1}$$

In addition, mark the user preference of users according to keywords, such as sports, science, and amusement. Generate user preference vector according to keywords and measure the hobby similarity of different users according to cosine similarity. The users with similar hobbies may transmit data. The cosine similarity $\mathrm{Sim}_{\mathrm{i,j}}$ between users i and j can be expressed as

$$\mathrm{Sim}_{i,j} = \frac{\overrightarrow{P_i} \cdot \overrightarrow{P_j}}{\|P_i\| \cdot \|P_j\|} \tag{2}$$

And Pi and Pj, respectively, represent the preference vectors between users i and j. The social connection strength between users i and j can be obtained according to the above formulas:

$$\mathrm{So}_{\mathrm{i,j}} = \beta \cdot \mathrm{F}_{\mathrm{i,j}} + (1 - \beta) \cdot \mathrm{Sim}_{\mathrm{i,j}} \tag{3}$$

And β is an adjustment factor. According to Shannon's theorem, the transmission rate $\mathrm{R_i}$ and BS between user i and base station is

$$\mathrm{R}_{\mathrm{i.BS}} = \mathrm{wlog}\left(1 + \frac{\mathrm{P_{bs}} \mathrm{d}_{\mathrm{bs}}^{-\alpha}}{\mathrm{No}}\right) \tag{4}$$

And w is bandwidth, Pbs is the transmitted power of base station, dbs is the distance between users and base station and α is decline factor. The transmission rate between users i and j is

$$R_{i,j} = wlog\left(1 + \frac{P_{i,j} d_{i,j}^{-\alpha}}{No}\right) \tag{5}$$

The quality of standardized physical link between users i and j is

$$R_{i,j}^{*} = \frac{R_{i,j} - \min\{R_j\}}{\max\{R_j\} - \min\{R_j\}} \tag{6}$$

And min{Rj} represents the lowest rate all content providers can achieve, max{Rj} represents the highest rate. At last, the utility between users i and j can be expressed as

$$\mathrm{u}_{\mathrm{i,j}} = \gamma \cdot \mathrm{So}_{\mathrm{i,j}} + (1 - \gamma) \cdot R_{i,j}^{*} \tag{7}$$

And γ is adjustment factor. Therefore, mode selection question can be expressed as the following optimization question:

$$max \sum\nolimits_{i=1}^{N} \sum\nolimits_{j=1}^{M} \alpha_{i,j} u_{i,j} \tag{8}$$

$$s.t. \sum_{j=1}^{M} \alpha_{i,j} \leq 1, \forall i \in [1, N] \tag{9}$$

And $\alpha_{i,j}$ is a binary variable. When the receiver i obtains data from the sender j, $\alpha_{i,j} = 1$, otherwise, $\alpha_{i,j} = 0$. The paper designs a random algorithm to solve the mode selection problem. First, initialize the state of user and calculate the quality of physical connection and strength of social relation among all users. All receivers are considered receiving content from base station. Second, set maximum iterations, randomly choose a receiver in each iteration and look for a connection plan improving user utility. At last, when the maximum iterations are achieved, end the algorithm. Mode selection algorithm is as follows.

Algorithm 1 mode selection

1. Initialization state information, maximum iterations round, initial pairing plan p, user relation matrix f, user preference matrix k
2. Calculate user relation closeness matrix F according to f
3. Calculate user similarity matrix S according to k
4. Calculate user utility matrix and user's current utility e according to F and S
5.for r=1: round
6. select a receiver i
7. for j=1:M
8. if utility(i,j)>e(i)
9. e(i)=utility(i,j)
10. p(i)=j
11. end if
12. end for
13.end for

4 Numerical Results

Assuming the radius of cellular district is 500 m, users randomly distribute in district, base station SNR = 2 dB, the maximum transmitted power of base station is 2 W, the maximum transmitted power of users is 0.126 W, noise power is 10^{-10} W, information channel obeys single path loss, loss coefficient $\alpha = 4$ and the specific simulation parameters are shown in Table 1.

Figure 2 describes the comparison of social-aware algorithm and non-social-aware algorithm, when the numbers of D2D users are respectively 100, 50, and 25. the non-

Table 1. Simulation parameters

Cell radius (m)	500
Base station SNR (dB)	2
Maximum transmit power of base station (W)	2
Maximum transmit power of user (W)	0.126
Noise power (W)	10^{-10}
Path loss index	4

social-aware algorithm is based on the connection rate of users. It can be found that as iterations increase, total user utility first improves gradually, then becomes stable. That is because users improve utility by changing pairing objects after iteration starts and utility finally reaches a state of balance with the constant optimization of pairing ways. The more users are, the slower the algorithm convergence is and the higher total user utility is. When the number of users is different, the total user utility of social-aware algorithm is obviously higher than non-social-aware algorithm indicating the social-aware-based mode selection algorithm can significantly improve user utility and social-aware algorithm has a more obvious advantage with the increase of the number of users. Meanwhile, it can be found through comparison that social-aware algorithm has a slightly slower convergence speed than non-social-aware algorithm.

Figure 3 illustrates the total user utility for the Social-Aware Mode Selection Algorithm (SAMSA) versus the HMSA and MSABoD algorithms [11]. The results

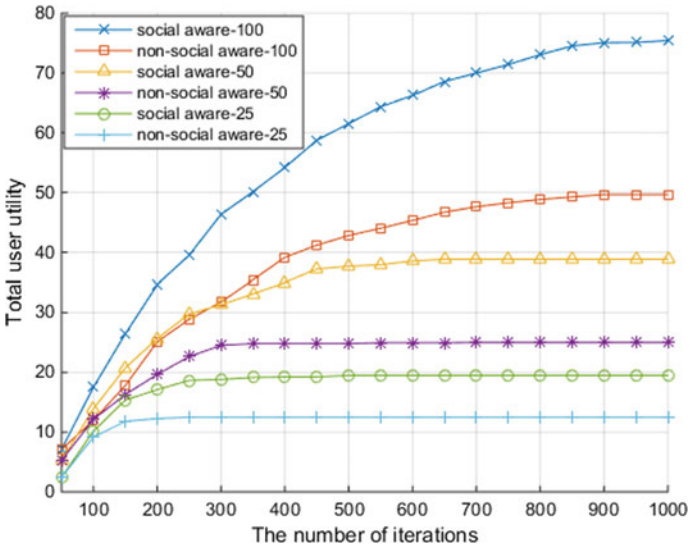


Fig. 2. The comparison of social-aware algorithm and non-social-aware algorithm when the number of users is different

show that, although the total utility of the three algorithms increases with the number of users, the total utility of SAMSA and HMSA users is significantly higher than that of MSABoD. This is because distance-based mode selection simply determines whether or not to use D2D communication based on distance, whereas distance does not fully reflect network status, which results in a lower overall rate of acquisition and MSABoD does not account for the user's social relationships ultimately leading to its user's low utility. However, the HMSA algorithm can obtain a higher total rate and make its user utility significantly higher than that of the MSABoD algorithm. However, the HMSA algorithm also does not consider the user's social relationship, so the total utility of the end user is still lower than that of the SAMSA algorithm. When the number of users is

100, the total user utility of the SAMSA algorithm is about 21% higher than that of the HMSA algorithm, which is about 72% higher than that of the MSABoD algorithm.

Figure 4 depicts the total rate of SAMSA versus HMSA and MSABoD algorithms. The results show that although the total rate of the three algorithms increases with the

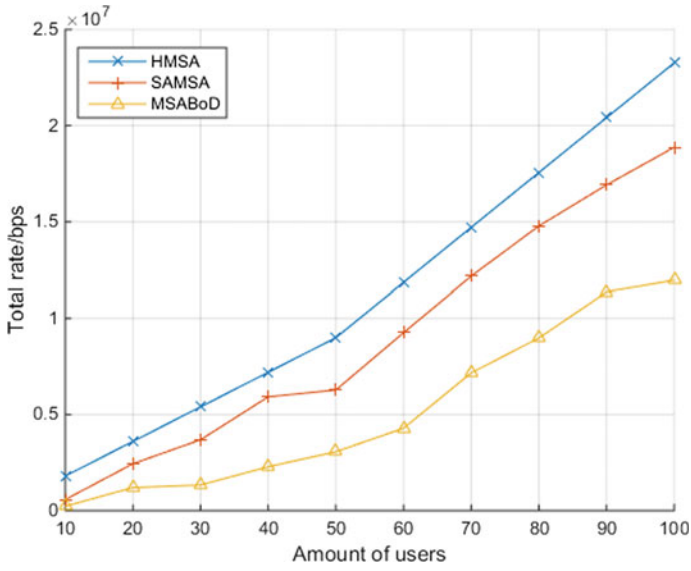


Fig. 3. Total rate comparison of social perception algorithm and heuristic algorithm

number of users, the total rate of the HMSA and SAMSA algorithms is always higher than that of the MSABoD algorithm. The lower MSABoD algorithm rate is due to its less efficient mode-based selection strategy. The SAMSA algorithm uses the transmission rate as one of the metrics and also considers the user's social relationship. Therefore, the mode selection does not pursue the maximum rate but the maximum user utility. The HMSA algorithm uses the rate as an optimization target and pursues the SINR And the maximum transmission power to meet the requirements of the maximum transmission rate, without considering the social relations of users, making the final total HMSA algorithm than the SAMSA algorithm. When the number of users is 100, the total rate of HMSA algorithm is about 28% higher than that of SAMSA algorithm, which is about 92% higher than that of MSABoD algorithm.

According to the comparison of Figs. 3 and 4, we can see that although the SAMSA algorithm can achieve high user utility, it cannot maximize the transmission rate. HMSA algorithm can achieve higher transmission rate but lower user utility than SAMSA algorithm. Therefore, the HMSA algorithm is suitable when the system scenario is pursuing the maximum transmission rate without regard to social relationships, and the SAMSA algorithm performs better when it comes to taking user social relationships into account.

Figure 5 describes the user utility of social-aware algorithm when β has different values. β decides the proportions of relational closeness and user similarity in user's

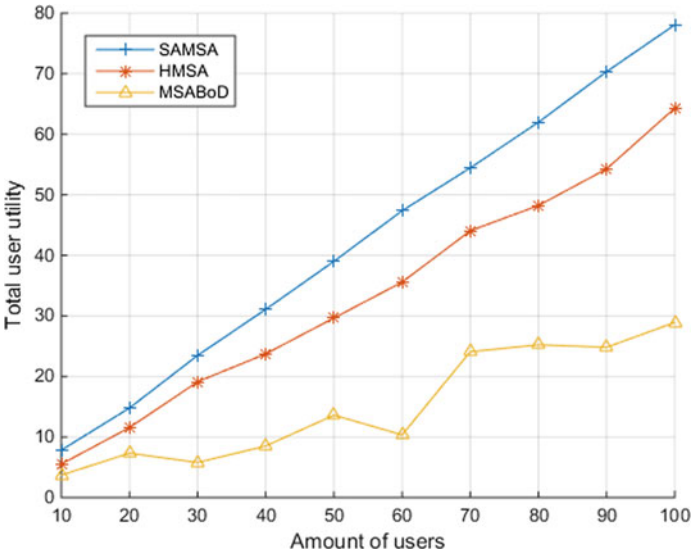


Fig. 4. Utility comparison of social-aware algorithm and heuristic algorithm

social relation. The larger β is, the higher the proportion relational closeness has. The fewer β is, the larger proportion user similarity has. Namely, as β increases, users tend to pair with the users closer to themselves and strange users have a low pairing probability, which sometimes stops some users outside relationship network; as β decreases, users tend to pair with the users with similar preference because the users with similar preference are more likely to own the data they need or transmit data to them in the future. However, if users have a low preference similarity, user utility is low. Because the user preference matrix in the simulation is sparse, namely users do not have a common preference, user utility lowers with the decrease of β.

Figure 6 describes the user utility of social-aware algorithm when γ has different values. γ decides the proportions of social relation and physical state in user utility. As γ increases, users tend to pair with the users closer to them. As γ decreases, users tend to pair with the users with a higher quality of the physical connection. Namely, as γ increases, users pay more attention to social relation. And as γ decreases, users pay more attention to the quality of the physical link. In the simulation, district communication has a high quality, base station can provide a good transmission rate and there is no social relation between the base station and users, which makes user utility decrease with the increase of γ. It indicates that when the physical link has a good quality and the social relation among users is not close, the quality of transmission should be valued. When there is a close relation between users and neighboring users, social relation should be valued.

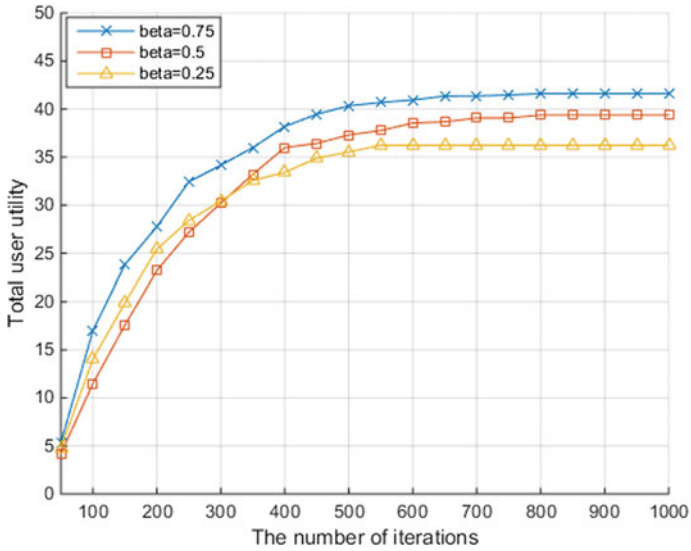


Fig. 5. The comparison of user utility of social-aware algorithm when β has different values

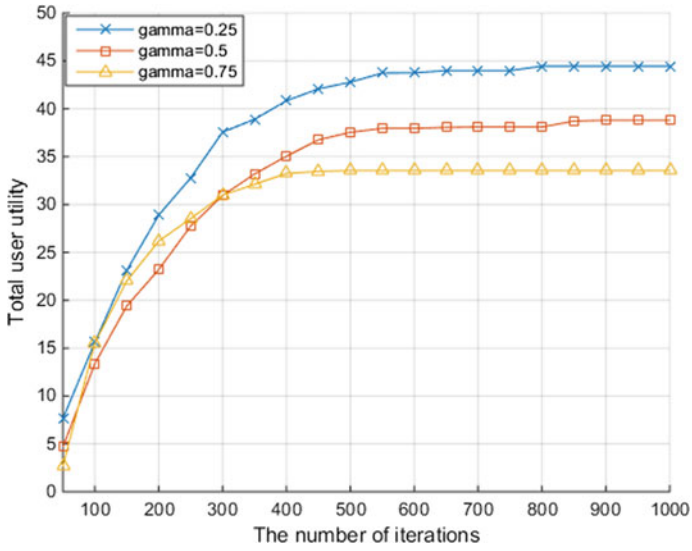


Fig. 6. The comparison of user utility of social-aware algorithm when y has different values

5 Conclusions

The paper studies a social-aware-based D2D communication mode selection strategy and designs a random matching algorithm which takes advantage of user similarity and relational closeness to measure user's social relation and ensures user utility by

introducing social aware in mode selection. The simulation results show that the introduction of social-aware can effectively improve user utility. And the proportions of user similarity, relational closeness, and quality of physical connection in user utility have important influences on the results of the algorithm.

Acknowledgements. *Foundation Items*: The National Natural Science Foundation of China for Youth (61201160, 61602263); The Natural Science Foundation of Jiangsu Province (BK20131377, BK20151507, BK20160916); The Natural science fund for colleges and universities in Jiangsu Province under Grants (16KJB510034); The six talent peaks project in Jiangsu Province (XYDXXJS-044); A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (yx002001); The Jiangsu Overseas Research and Training Program for University Prominent Young and Middle-aged Teachers and Presidents; Sponsored by NUPTSF (Grant Nos. NY212012, NY214065,NY216020)

References

1. Pavel, M., Zdenek, B., Tomas, V.: In-band device-to-device communication in OFDMA cellular networks: a survey and challenges. IEEE Commun. Surv. Tutor. **17**(4), 1885–1922 (2015)
2. Li, Y., Gursoy, M.C., Velipasalar, S.: A delay-aware caching algorithm for wireless D2D caching networks: In: IEEE Conference on Computer Communications Workshops (2017)
3. Xu, Y.A.: Mode selection scheme for D2D communication in heterogeneous cellular networks. In: 2015 IEEE Global Communications Conference. United States: Institute of Electrical and Electronics Engineers Inc, San Diego (2015)
4. Ahmed, E., Yaqoob, I., Gani, A., et al.: Social-aware resource allocation and optimization for D2D communication. IEEE Wirel. Commun. **24**(3), 122–129 (2017)
5. Cai, Y., Wu, D., Yang, W.: Social-aware content downloading mode selection for D2D communications. In: IEEE International Conference on Communications, pp. 2931–2936. Institute of Electrical and Electronics Engineers Inc, London (2015)
6. Wu, D., Zhou, L., Cai, Y.: Social-aware distributed joint mode selection and link allocation for mobile D2D communications. In: 2017 IEEE International Conference on Communications Workshops, pp. 851–856. Institute of Electrical and Electronics Engineers Inc, Paris (2017)
7. Wu, D., Zhou, L., Cai, Y.: Social-aware rate based content sharing mode selection for D2D content sharing scenarios. IEEE Trans. Multimed. **19**(11), 2571–2582 (2017)
8. Zhao, Y., Li, Y., Cao, Y., et al.: Social-aware resource allocation for device-to-device communications underlaying cellular networks. IEEE Trans. Wireless Commun. **14**(12), 6621–6634 (2015)
9. Cao, Y., Jiang, T., Chen, X., et al.: Social-aware video multicast based on device-to-device communications. IEEE Trans. Mob. Comput. **15**(6), 1528–1539 (2016)
10. Yang, L., Wu, D., Xu, S., et al.: Social-aware energy-efficient joint mode selection and link allocation in D2D communications. In: 2016 8th International Conference on Wireless Communications and Signal Processing. Institute of Electrical and Electronics Engineers Inc, Yangzhou (2016)
11. Lu, W., Lin, W., Yang, L., et al.: A heuristic D2D communication mode selection algorithm. In: 2017 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (2017)

Joint Pilot and Power Assignment to Improve the Fairness of Massive MIMO System

Shuangshuang Jiang(✉) and Bin Wang

School of Electronics and Information Engineering,
Harbin Institute of Technology, Harbin 150001, China
16s105147@stu.hit.edu.cn

Abstract. In massive MIMO system, each user has different channel quality, so this paper considers the aspect of improving the system fairness. It carries out both the power assignment and the pilot assignment to improve the spectral efficiency of users with poor channel quality as much as possible. The use of TDD mode in a multi-cell system is considered, and in the uplink two receiver technologies which are maximum-ratio combining (MRC) and zero-forcing (ZF) receiver will be analyzed, respectively. Applying the lower bound of the spectral efficiency to optimize the pilot and data power then joint pilot assignment at the same time, an optimization model to maximize the minimum spectral efficiency which can improve the fairness of the system is proposed. Simulation results verify the effectiveness of the proposed joint assignment scheme.

Keywords: Massive MIMO · Power assignment · Joint assignment · System fairness

1 Introduction

Due to the high capacity and high reliability of the massive MIMO system, it is widely studied in 5G technology. Since a large number of antennas are installed at the base station, the TDD mode has been widely considered in the massive MIMO system. In this mode, the base station processes the data using the received pilot signals in the uplink [3]. Because the same pilot sequences in different cells will generate interference, which is called pilot contamination, effective pilot and power assignment are necessary especially when to improve system fairness.

There are many literature about resource assignment to improve system performance. In [4], the authors used different linear receiver techniques in the uplink and carried on a detailed analysis of system performance using different receiver technologies. In [5], the authors carried out pilot and data power assignment for the proposed optimization model which maximizes sum spectral efficiency and finds that more power should be allocated to pilot transmission when channel conditions are poor. A SPA (smart pilot assignment) algorithm was proposed to improve the fairness of the system and improve the SINR of users with poor channel quality in [6]. The MMSE receiver in the uplink was considered and minimized the total pilot and data power under the constraint conditions in [1]. In [2], the authors proposed an optimization

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 725–732, 2019.
https://doi.org/10.1007/978-981-13-6264-4_85

model that minimizes the correlation channel error considering the total pilot power assignment which effectively reduces the channel error.

This paper considers that the uplink applies MRC and ZF receivers, respectively, and proposes an optimization model that maximizes the minimum spectral efficiency which applies the lower bound of spectral efficiency expressions with a joint pilot assignment and power assignment. It optimizes pilot and data power in the case of fixed pilot assignment. Using the obtained pilot and data power combines pilot assignment to further optimize the objective function. The simulation results show that the proposed joint assignment method can significantly improve the minimum spectral efficiency and prove the effectiveness of the proposed optimization model.

2 System Model

This paper considers a massive MIMO system with L cells and the system uses TDD mode. The number of base station's antennas is M. There are K users in each cell. The orthogonal pilot sequences are used in one cell. Different cells reuse this orthogonal pilot sequences [3] (Fig. 1).

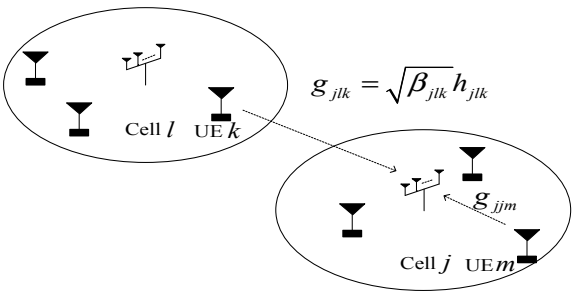


Fig. 1. Multi-cell system model

2.1 Pilot Transmission

In this stage, users send pilot information to the base station j. The pilot length τ is larger than K, and the receiver matrix obtained by the base station can be expressed as

$$Y_j^{pilot} = \sum_{i=1}^{L} \sum_{t=1}^{K} \sqrt{\tau p^{pilot} \beta_{jit}} h_{jit} \psi_{it} + N_j^{pilot} \tag{1}$$

where N_j^{pilot} denotes normalized noise, p^{pilot} denotes pilot power, Ψ_{it} denotes pilot sequence used by user t in cell i, $\sqrt{\beta_{jit}} h_{jit}$ is expressed as the channel information g_{jit} of user t in cell i to cell j, h_{jit} denotes fast fading coefficient and $\sqrt{\beta_{jit}}$ denotes slow fading coefficient.

The base station applies MMSE channel estimation and the channel estimation formula of user k in cell j to cell j can be expressed as

$$\hat{g}_{jjk} = \frac{\sqrt{\tau p^{pilot}}\beta_{jjk}}{1 + \sum_{l=1}^{L} \tau p^{pilot}\beta_{jlk}} Y_j^{pilot} \psi_{jk}^H \tag{2}$$

MMSE channel estimation error is $\tilde{g}_{jjk} = g_{jjk} - \hat{g}_{jjk}$, we get $\hat{g}_{jjk} \sim \mathcal{CN}\left(0, \sigma_{jjk}^2 I_M\right)$ and $\tilde{g}_{jjk} \sim \mathcal{CN}\left(0, \varepsilon_{jjk}^2 I_M\right)$ with

$$\sigma_{jjk}^2 = \frac{\tau p^{pilot}\beta_{jjk}^2}{1 + \sum_{l=1}^{L} \tau p^{pilot}\beta_{jlk}} \tag{3}$$

$$\varepsilon_{jjk}^2 = \beta_{jjk} - \frac{\tau p^{pilot}\beta_{jjk}^2}{1 + \sum_{l=1}^{L} \tau p^{pilot}\beta_{jlk}} \tag{4}$$

As seen from the above expression (2), users using the same pilot sequences between different cells interfere with channel estimation [2].

2.2 Data Transmission

At this stage, the base station receives data information sent by all users, and the information received by the base station j can be expressed as

$$y_j^{data} = \sqrt{p^{data}} \sum_{i=1}^{L} \sum_{t=1}^{K} g_{jit} q_{it} + N_j^{data} \tag{5}$$

where N_j^{data} denotes normalized noise, p^{data} denotes data power, q_{it} denotes data information sent by user t in cell i.

The base station j multiplies the linear detection matrix w_{jjk}^H by the received information y_j^{data}. Then the expression can be obtained as

$$\begin{aligned} \bar{y}_{jjk}^{data} &= w_{jjk}^H y_j^{data} \\ &= \left(\sqrt{p^{data}} w_{jjk}^H \hat{g}_{jjk} q_{jk} + \sqrt{p^{data}} \sum_{i=1}^{L} \sum_{t=1, t\neq k}^{K} w_{jjk}^H \hat{g}_{jit} q_{it} + w_{jjk}^H N_j^{data}\right. \\ &\quad \left. + \sqrt{p^{data}} \sum_{i=1, i\neq j}^{L} w_{jjk}^H \hat{g}_{jik} q_{ik} + \sqrt{p^{data}} \sum_{i=1}^{L} \sum_{t=1}^{K} w_{jjk}^H \tilde{g}_{jit} q_{it}\right) \end{aligned} \tag{6}$$

where for MRC and ZF receivers, respectively, w_{jjk}^H can be expressed as [4]

$$w_{jjk} = \begin{cases} \hat{g}_{jjk} & MRC \\ \hat{g}_{jjk}\left(\hat{g}_{jjk}^H \hat{g}_{jjk}\right)^{-1} & ZF \end{cases} \tag{7}$$

2.3 The Spectral Efficiency with MRC Receiver

When MRC receiver is used, there is $w_{jjk} = \hat{g}_{jjk}$, applying Jensen's inequality, the lower bound of spectral efficiency for user k in cell j can be expressed as [5]

$$S_{jk}^{MRC} = \frac{T-\tau}{T}\log_2\left(1 + \frac{\tau p^{pilot} p^{data}(M-1)\beta_{jjk}^2}{\tau p^{pilot} p^{data}\left(\theta_{jk}\right) + p^{data}\left(\chi_{jk}\right) + \tau p^{pilot}\sum\limits_{l=1}^{L}\beta_{jlk} + 1}\right) \tag{8}$$

where $\theta_{jk} = (M-1)\sum\limits_{i=1,i\neq j}^{L}\beta_{jik}^2 - \sum\limits_{i=1}^{L}\beta_{jik}^2 + \sum\limits_{l=1}^{L}\beta_{jlk}\sum\limits_{i=1}^{L}\sum\limits_{t=1}^{K}\beta_{jit}$, $\chi_{jk} = \sum\limits_{i=1}^{L}\sum\limits_{t=1}^{K}\beta_{jit}$, T denotes coherence time length.

2.4 The Spectral Efficiency with ZF Receiver

When ZF receiver is used, there is $w_{jjk} = \hat{g}_{jjk}\left(\hat{g}_{jjk}^H \hat{g}_{jjk}\right)^{-1}$, applying Jensen's inequality, the lower bound of spectral efficiency for user k in cell j can be expressed [4]

$$S_{jk}^{ZF} = \frac{T-\tau}{T}\log_2\left(1 + \frac{p^{data}(M-K)\sigma_{jjk}^2}{p^{data}\sum\limits_{i=1}^{L}\sum\limits_{t=1}^{K}\varepsilon_{jit}^2 + p^{data}(M-K)\sum\limits_{i=1,i\neq j}^{L}\sigma_{jik}^2 + 1}\right) \tag{9}$$

It can be seen that while keeping other parameters unchanged, the spectral efficiency is affected by the uplink pilot power, data power and pilot assignment conditions. A reasonable power assignment and simultaneously joint pilot assignment will effectively improve the system performance.

3 Maximizing the Minimum Spectral Efficiency

In the massive MIMO system, considering the fairness of users, the spectral efficiency of users with the worst channel quality of the system should be improved as much as possible. In this section, carrying out power assignment and simultaneously jointing pilot assignment, this paper proposes an optimization model that maximizes the minimum spectral efficiency, which can be expressed as

$$\begin{aligned} \underset{p^{pilot},p^{data},\eta_j}{\text{maximize}} \quad & min\left(S_{jk}\right), \forall j,k \\ s.t. \quad & \tau p^{pilot} + (T-\tau)p^{data} \le E \\ & p^{pilot} > 0, p^{data} > 0 \\ & \eta_j \in \{1, 2 \cdots K\}, \forall j \end{aligned} \tag{10}$$

In the above optimization model, η_j denotes the pilot assignment of all users in cell j. E denotes the uplink power limit of each user which can be proved that objective function is maximized when the equation is established. The pilot length τ is selected as the number of users in the cell which is the optimal pilot length [5]. The above optimization problem can be transformed into

$$\begin{aligned} \underset{p^{data},\eta_j}{\text{maximize}} \quad & min\left(S_{jk}\right), \forall j,k \\ s.t. \quad & p^{pilot} = \left(E - (T-K)p^{data}\right)/K \\ & 0 \le p^{data} \le \tfrac{E}{T-K} \\ & \eta_j \in \{1, 2 \cdots K\}, \forall j \end{aligned} \tag{11}$$

It can be proved that under the given assignment of pilots, the above objective function is a concave function [5], so there is an optimal solution to the power solution. In the case of solving the optimal power assignment, the joint SPA method is used for pilot assignment at the same time [6].

Given the pilot assignment, the fmincon function which is the nonlinear programming solver in optimization toolbox of MATLAB is used to solve Problem (11) to get the optimal data power. Applying the obtained data power, the channel quality of the user in the current cell can be expressed as ξ_k, where the subscript k denotes the user k. The interference caused by users in neighboring cells using the same pilot can be expressed as ς_k, where the subscript k denotes the pilot k. Sort the user's channel quality and the interference in ascending order, respectively. Assigning pilot sequences with low interference to users with poor channel quality, successively. Finally, iteratively find the optimal solution. The solving algorithm is shown in Table 1.

4 Numerical Results

The simulation considers $L = 7$ cells, the number of base station's antennas is $M = 100$, the number of users in every cell is $K = 10$, and cell's radius is 1000 m. All users are uniformly distributed within the area of $r_h = 200$ m away from the origin of the base station. The coherence time of the channel is $T = 200$. Large-scale fading can be expressed as $\beta_{jik} = z_{jik}/(r_{jik}/r_h)^v$, where $10\log10(z_{jik})$ obeys the Gaussian distribution, the standard deviation is 8 dB, and $v = 3.8$ is the path loss index, assuming that $E_0 = E/T$ represents the average transmit power.

According to the above-mentioned large-scale fading model, 2000 times simulations are generated randomly. The cumulative distribution function (CDF) curves of

Table 1. Maximize the minimum spectral efficiency

Input: Number of iterations N, E, β_{jik}, p^{pilot}, p^{data}
1: Initialize pilot assignment: assign pilot Ψ_k to user k
2: for n = 1: N
3: apply fmincon function to solve (13) to get (p^{pilot}, p^{data}, S_{jk})
4: for j = 1:L
5: for k = 1:K
6: $\xi_k = p^{data}(E-(T-K)p^{data})\beta_{jjk}^2$; $\varsigma_k = \sum_{i=1,i\neq j}^{L} p^{data}(E-(T-K)p^{data})\beta_{jik}^2$;
7: end
8: *ascending sort*($[\xi_1, \xi_2, \cdots \xi_K]$); *ascending sort*($[\varsigma_1, \varsigma_2, \cdots \varsigma_K]$);
9: assign pilot Ψ_k to user k
10: end
11: end
Output: p^{pilot}, p^{data}, S_{jk}

minimum uplink SINR using MRC receiver are shown in Fig. 2(a) and (b). The CDF of minimum uplink SINR is computed under the condition of the average transmit power E_0 = 0 dB (as shown in Fig. 2(a)) and E_0 = 10 dB (as shown in Fig. 2(b)). The simulation results illustrate that the performance of the joint assignment method significantly increases the system's minimum uplink SINR. Under the condition of E_0 = 0 dB, with the aid of the MRC receiver, the minimum uplink SINR reached by at least 95% of the users is −70.5 dB without assignment. After the joint assignment is performed, the minimum uplink SINR achieves −60.5 dB. Taking advantage of the proposed joint assignment method, a gain of more than 10 dB is obtained in the minimum uplink SINR. According to the simulation results shown in Fig. 2(a) and (b), observe the performance gain brought by the pilot assignment, it can be found that the

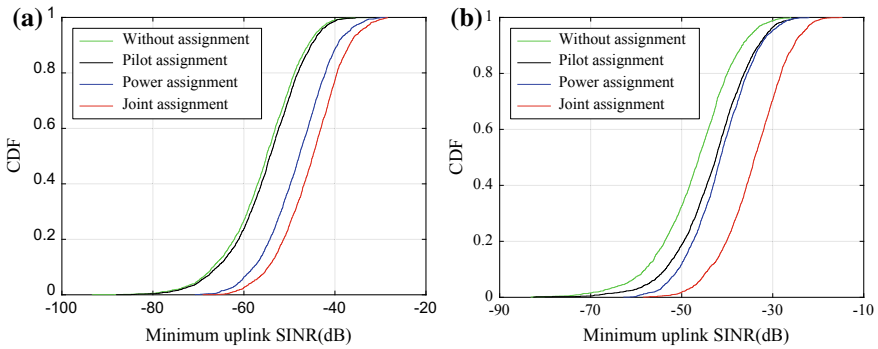


Fig. 2. CDF of minimum uplink SINR with MRC receiver (a) Minimum uplink SINR (E_0 = 0 dB). (b) Minimum uplink SINR (E_0 = 10 dB)

performance gain is more obvious when the average transmission power is high. Then observe the performance gain caused by the power assignment, it is more significant when the average transmission power is low. In Fig. 2(a) and (b), we can find that the joint assignment can significantly increase the minimum uplink SINR.

The CDF curves of minimum uplink SINR using ZF receiver are shown in Fig. 3(a) and (b). The simulation results are computed under the condition of the average transmit power E_0 = 0 dB (as shown in Fig. 3(a)) and E_0 = 10 dB (as shown in Fig. 3 (b)). It can be analyzed in the same way as above, when using ZF receiver, considering that the minimum uplink SINR reached by at least 95% of the users, after the joint assignment, more than 10 dB gain can be improved compared without assignment. Performing joint assignment significantly increases the minimum uplink SINR.

Considering the average transmit power E_0 = 0 dB, compare the CDF curve of spectral efficiency of each user before and after performing joint assignment with MRC and ZF receivers, respectively, as shown in Fig. 4. It can be seen that when the user's spectral efficiency is lower, the two receivers are basically the same. With per user's spectral efficiency increasing, compared with applying MRC receiver, applying ZF receiver makes the user with the same probability achieve higher spectral efficiency, and it can be found that applying ZF receiver has a greater advantage in high SINR. In addition, it can be seen that when the spectral efficiency of each user is low, the minimum value of the user spectral efficiency is improved and the spectral efficiency is more concentrated after the joint assignment. The simulation result shows that the system fairness is improved.

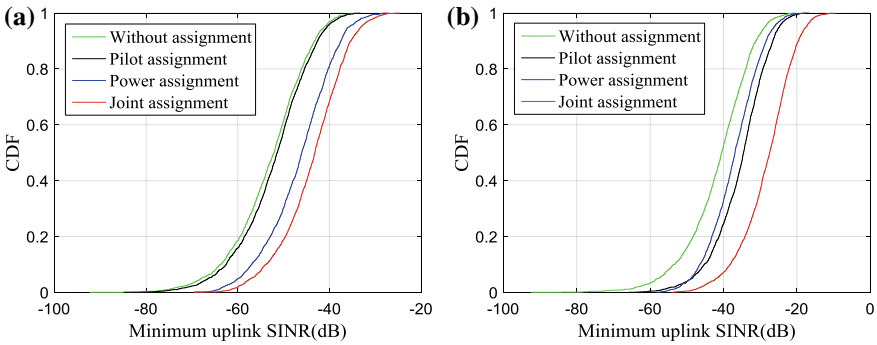


Fig. 3. CDF of minimum uplink SINR with ZF receiver (a) Minimum uplink SINR (E_0 = 0 dB). (b) Minimum uplink SINR (E_0 = 10 dB)

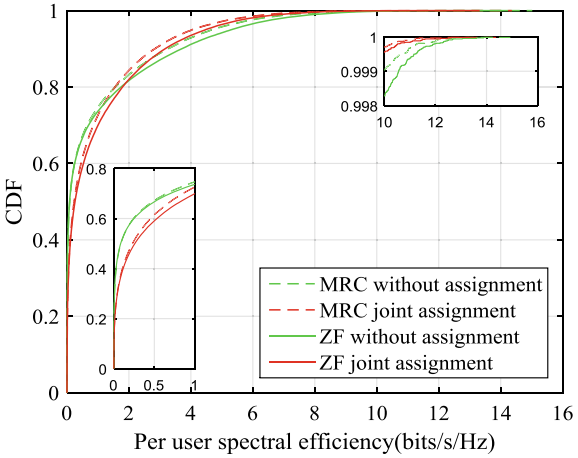


Fig. 4. CDF of per user spectral efficiency with MRC and ZF receiver, respectively (E_0 = 0 dB)

5 Conclusion

In this paper, considering the uplink applying MRC and ZF receiver, respectively, the optimization model that maximizes the minimum spectral efficiency has been proposed which considers power assignment between pilot and data power then joints pilot assignment at the same time. In the case of a given pilot assignment, the objective function has the optimal solution of the power assignment and the obtained power result is combined with the pilot assignment to further optimize the objective function. The simulation results showed that considering the performance gain caused by the pilot assignment, it is more obvious when the average transmission power is high. Then observing the performance gain caused by the power assignment, it is more significant when the average transmission power is low. And the joint assignment can significantly improve the system fairness.

References

1. Guo, K.F., Guo, Y., Fodor, G.B., et al.: Uplink power control with MMSE receiver in multi-cell MU-massive-MIMO systems. In: ICC 2014, pp. 5184–5190 (2014)
2. Liu, P., Jin, S., Jiang, T., et al.: Pilot power allocation through user grouping in multi-cell massive MIMO systems. IEEE Trans. Commun. **65**, 1561–1574 (2017)
3. Marzetta, T.L.: Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans. Wireless Commun. **9**, 3590–3600 (2010)
4. Ngo, H.Q., Larsson, E.G., Marzetta, T.L.: Energy and spectral efficiency of very large multiuser MIMO systems. IEEE Trans. Commun. **61**, 1436–1449 (2013)
5. Ngo, H.Q., Matthaiou, M., Larsson, E.G.: Massive MIMO with optimal power and training duration allocation. IEEE Commun. Lett. **3**, 605–608 (2014)
6. Zhu, X.D., Wang, Z.C., Dai, L.L., et al.: Smart pilot assignment for massive MIMO. IEEE Commun. Lett. **19**, 1644–1647 (2015)

Satellite Autonomous Integrity Monitoring (SAIM) for Satellite Clock Slow Anomaly

Yizhe Jia(✉), Lang Bian, and Lixin Zhang

China Academy of Space Technology (Xi'an), 504 Chang'an Avenue, 710100 Xi An, China
helios_yizhe@qq.com

Abstract. There are mainly two types of navigation satellite clock in-orbit failures, namely fast anomaly and slow anomaly. The fast anomaly refers to a sudden change in clock error, while the slow anomaly refers to a clock bias that slowly deviates from normal. At present, fast anomaly has been effectively monitored in-orbit, and slow anomaly has problems with long latency and is susceptible to fast anomaly, and has not been effectively solved. This paper analyzes the characteristics of the two types of anomaly, and studies the satellite autonomous integrity monitoring method for satellite clock slow anomaly that has not been effectively solved in orbit. A method for satellite autonomous integrity monitoring based on least-squares quadratic polynomial fitting and separation of fast anomaly and slow anomaly is proposed. Slow anomaly monitoring was implemented based on the separation of two types of failures. Simulation experiments verify the feasibility of this method and evaluate the time-to-alert.

Keywords: SAIM · Navigation satellite clock · Slow anomaly · Time-to-alert

1 Introduction

With the increasing use of GNSS, the demand for the integrity of navigation satellites is also increasing. Satellite clock is the core component of navigation satellites. Its integrity service performance is an important part of the integrity of navigation satellites [1]. Integrity monitoring of navigation satellite clocks can be divided into three categories, namely, receiver autonomous integrity monitoring (RAIM), augmentation system integrity monitoring, and satellite autonomous integrity monitoring (SAIM). The first two ground-based monitoring methods can reduce satellite payload complexity and reduce satellite costs. But ground monitoring cannot cover and monitor all navigation satellites in real time. On the one hand, RAIM needs at least five satellites to detect faults, and at least six satellites are needed to identify and eliminate faults. There are high requirements for satellite constellation configuration. On the other hand, the transmission of signals over many satellites will significantly affect the time-to-alert (TTA) and have poor alarm capability in time. With increasingly stringent integrity requirements [2], the pressure and risk of ground-based monitoring increase. The integrity monitoring of the satellite clock in the space segment by SAIM can shorten the TTA, reduce the risk, and improve the GNSS integrity service performance.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 733–741, 2019.
https://doi.org/10.1007/978-981-13-6264-4_86

Improving the performance of integrity services is also an inevitable requirement for the future development of GNSS [3, 4].

The navigation satellite clock in-orbit malfunctions can be divided into fast anomaly and slow anomaly. In the fast anomaly, article [5] proposed a method for autonomous integrity monitoring of atomic clocks, but did not take into account the slow anomaly of satellite clocks. Article [6] proposed a SAIM method based on Galileo inter-satellite link, but no detailed analysis was performed for slow anomaly, and the latent period of slow anomaly was long. Article [7] designed the SAIM scheme based on the BDS, but mainly related to satellite signal quality and satellite clock fast anomaly monitoring. There was no in-depth study on satellite clock slow anomaly monitoring. At present, fast anomaly has been effectively solved and in-orbit verification has been conducted [8]. Slow anomaly has long fault latency and is susceptible to fast anomaly. At present, due to the existence of certain in-orbit monitoring difficulties, it has not been effectively monitored in orbit.

This paper analyzes the fast-anomaly and slow-anomaly characteristics of navigation satellite clocks, and proposes a SAIM method that can separate two types of malfunctions from navigation satellite clocks. The algorithm is verified by using the measured clock bias data. After the simulation analysis, the method proposed in this paper can separate the fast anomaly from the slow anomaly. While monitoring slow anomaly, false alarms caused by fast anomaly are avoided.

2 Classifications of Clock Anomalies

Stanford University statistics recorded the GPS satellite signal-in-space anomalies from January 2004 to August 2010. There were 23 occurrences of satellite clock anomalies and 7 occurrences of ephemeris anomalies [9]. According to the analysis, the navigation satellite clock and ephemeris faults are the most frequent forms of satellite in-orbit failures, of which satellite clocks have the highest frequency of failures. The satellite clock anomaly is mainly divided into two aspects [10]:

(1) Fast anomaly: Fast anomaly can also be divided into jump anomaly and step anomaly. Among them, jump anomaly means that the satellite clock bias jumps at a certain moment and returns to normal after the next moment or a few short moments. And step anomaly refers to that the satellite clock bias jumps at a certain moment and does not return to normal afterward.
(2) Slow anomaly: Satellite clock bias deviate from normal values at a certain rate. The performance is not obvious at the time of fault occurrence, but it will accumulate a large clock error as time increases.

These anomalous effects may happen singly or in combination, suddenly, or over a period of time. Therefore, continuous monitoring of satellite clock is required. At present, fast anomaly has achieved integrity monitoring in orbit. Slow anomaly due to long latency and is susceptible to fast-anomaly disturbances. Therefore, the current satellite clock slow anomaly has not been effectively monitored in orbit.

3 Navigation Satellite Clock SAIM Algorithm

3.1 Quadratic Polynomial Fitting Clock Bias

In the GNSS broadcast ephemeris, the satellite clock bias can be represented by a quadratic polynomial, so that

$$\Delta T = a_0 + a_1 t + a_2 t^2/2 \tag{1}$$

where a_0 is the clock deviation, a_1 is the clock drift, and a_2 is the drift rate. The estimate of the clock bias can be calculated by fitting a quadratic polynomial with the least square method. Taking into account the load of the navigation satellite processor, the algorithm should have the characteristics of low computational complexity. Therefore, the following method is used when calculating polynomial coefficients.

For data points $(x_i, y_i), i = (0, 1, 2, \ldots, N-1)$, let Ψ be a set of functions consisting of polynomials of all times no more than two. The quadratic polynomial $p(x) = \sum_{k=0}^{2} a_k x^k \in \Psi$ make the following function take the minimum value, that is

$$I = \sum_{i=0}^{N-1} [p(x_i) - y_i]^2 = \sum_{i=0}^{N-1} \left(\sum_{k=0}^{2} a_k x_i^k - y_i \right)^2 \tag{2}$$

where $a_k (k = 0, 1, 2)$ are the coefficients of Eq. (1). Therefore,

$$\frac{\partial I}{\partial a_j} = 2 \sum_{i=0}^{N-1} \left(\sum_{k=0}^{2} a_k x_i^k - y_i \right) x_i^j = 0 \quad j = 0, 1, 2 \tag{3}$$

then we have the linear equations, so that

$$\sum_{k=0}^{2} \left(\sum_{i=0}^{N-1} x_i^{j+k} \right) a_k = \sum_{i=0}^{N-1} x_i^j y_i \quad j = 0, 1, 2 \tag{4}$$

The linear equations of (4) are represented by a matrix

$$\begin{bmatrix} N & \sum_{i=0}^{N-1} x_i & \sum_{i=0}^{N-1} x_i^2 \\ \sum_{i=0}^{N-1} x_i & \sum_{i=0}^{N-1} x_i^2 & \sum_{i=0}^{N-1} x_i^3 \\ \sum_{i=0}^{N-1} x_i^2 & \sum_{i=0}^{N-1} x_i^3 & \sum_{i=0}^{N-1} x_i^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{N-1} y_i \\ \sum_{i=0}^{N-1} x_i y_i \\ \sum_{i=0}^{N-1} x_i^2 y_i \end{bmatrix} \tag{5}$$

where the coefficient matrix of (5) is a symmetric positive definite matrix, so there is a unique solution. Thus, the three coefficients of the quadratic polynomial are obtained.

3.2 Monitoring Quantity Calculation

As (1) shows, if there is an abnormal change in the clock bias, the first-order and second-order coefficients will change. For example, a fast anomaly in clock bias can cause a_1 and a_2 jumps. Therefore, monitoring the first- and second-order coefficient changes to achieve separation and monitoring of fast and slowly anomaly of satellite clock. Difference between the first-order coefficients at any two adjacent moments to remove temperature-induced drift, i.e., $a_1(k) - a_1(k-1) = \Delta a_1$.

The quadratic polynomial is used to predict the clock bias at the next moment, and the measured value is subtracted to obtain the prediction error, which is denoted as *PreE*.

The window length of the least-squares fit selection is N, and no matter how abnormal the monitored quantity is, the duration is at most N seconds (one data per second). The difference between the $a_1(k)$ and $a_1(k-N)$ is denoted by $\Delta N a_1$. This can ensure that some small changes can be monitored.

3.3 The Logic of Monitoring

The four monitoring quantities used are *PreE,* a_2, Δa_1 and $\Delta N a_1$. The monitoring method is divided into two parts. The first part is the Step false alarm monitoring, as shown in Fig. 1. The second part is the slow-anomaly monitoring as shown in Fig. 2.

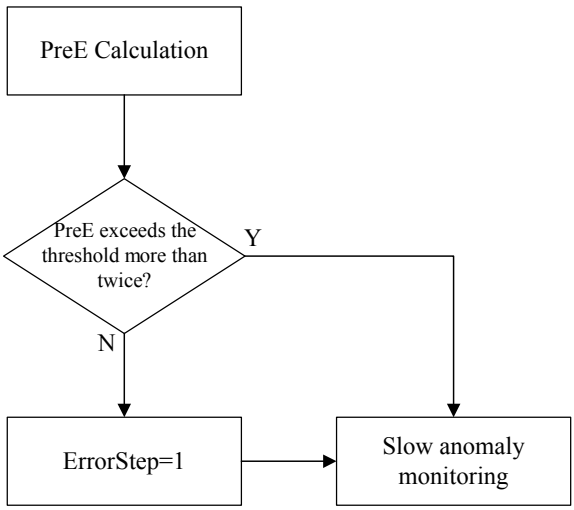


Fig. 1. Step anomaly false alarm monitoring

The main purpose of Fig. 1 is to rule out the influence of step anomaly on slow-anomaly monitoring. If *PreE* continuously exceeds the threshold for two epochs, we believe that there may be a slow anomaly. If the ErrorStep flag is 1, it proves that the fault is a step false alarm, and then the alarm is not performed in the slow-anomaly monitoring.

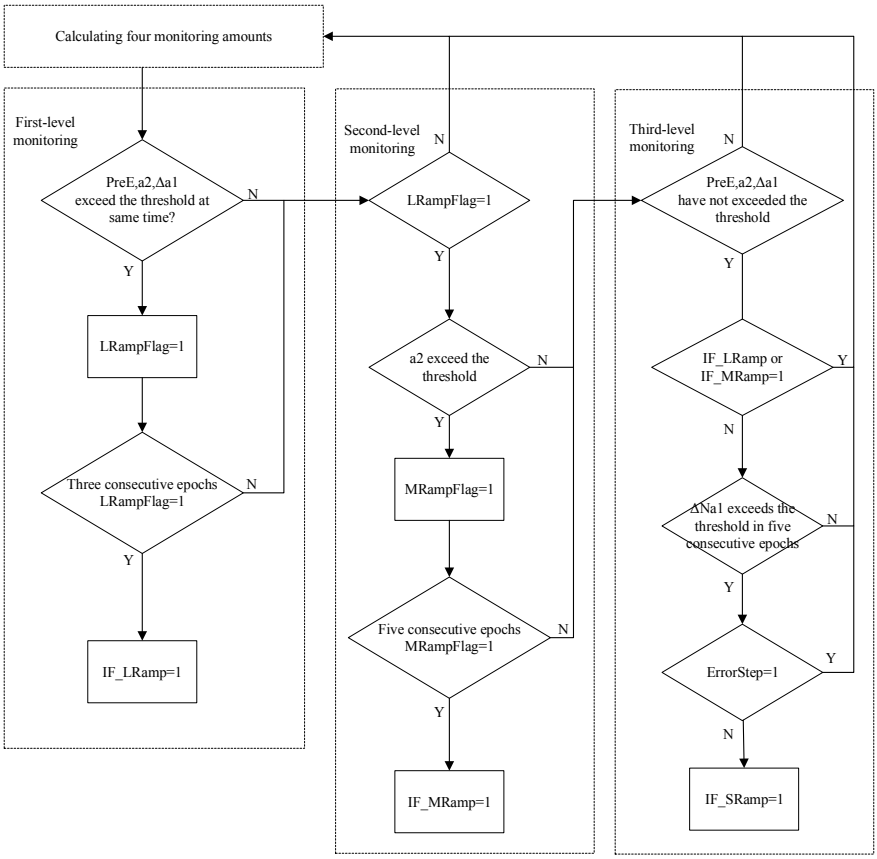


Fig. 2. Slow anomaly monitoring logic

The slow-anomaly monitoring in Fig. 2 is divided into three levels, mainly for different rates, thus ensuring the TTA to some extent. The data sequentially enters three levels of monitoring. If the flags (LRampFlag, MRampFlag) in a certain level are not generated continuously, they are set to zero.

4 Simulation Results

The data is measured by the clock bias of the navigation satellite without failure. The SAIM algorithm was failure-tested by injecting a clock bias ramp error. With a Gaussian model a probability of 10^{-7}, as required for category I precision approach (CAT-I), is reached by allowing data within 5.33 standard deviations. Multiplying the standard deviation by 5.33 to yield the threshold.

4.1 Slow Anomaly Monitoring

Injecting a 0.4 ns/s clock ramp error, the first 200 data are used to generate the monitoring threshold. The monitoring results are shown in Figs. 3, 4, and 5.

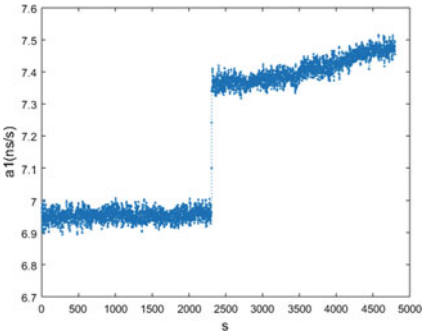


Fig. 3. a_1 Anomaly monitoring

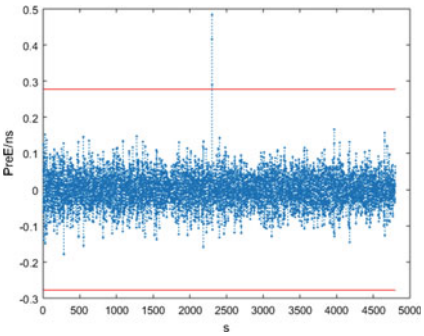


Fig. 4. *PreE* monitoring results

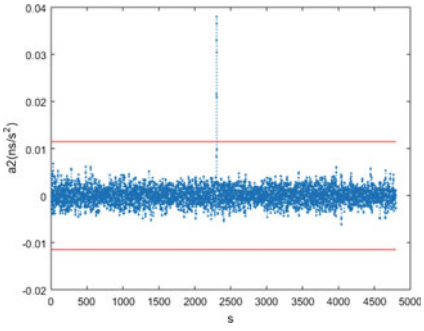


Fig. 5. a_2 anomaly monitoring

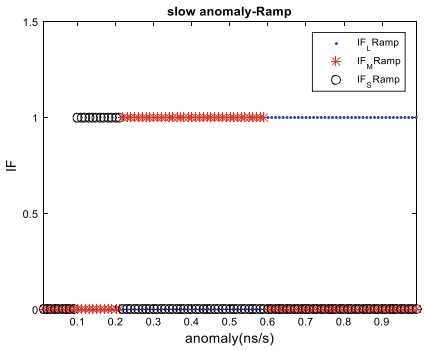


Fig. 6. Ramp error monitoring results

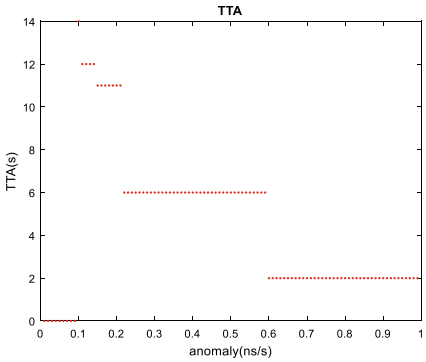


Fig. 7. The results of TTA

The monitoring results with different ramp error are shown in Fig. 6. When the ramp error is less than 0.1 ns/s, the rate is too small to be monitored. The TTA are shown in Fig. 7. The three-level monitoring can be clearly seen from the TTA. The TTA for monitoring at the first level is longer because the ramp error is too small and needs to be increased to a certain level to exceed the monitoring threshold to generate the integrity flag. Although the TTA is long, due to the slow ramp, no excessive errors will occur. The second and third levels of TTA are 6 s and 2 s, satisfying the TTA requirement of CAT-I, and further proves that the algorithm realizes the requirement of monitoring slow anomaly.

4.2 Fast Anomaly Separating

Assume that the fast-anomaly detection threshold is 1 ns, so the threshold for fault separation is 1 ns (i.e., slow-anomaly monitoring algorithm should not alert fast anomaly less than 1 ns). Therefore, a 0.01–0.99 ns clock fast anomaly was injected to simulate the algorithm. The results are shown in Figs. 8 and 9.

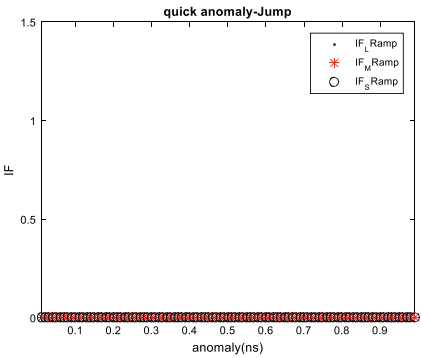


Fig. 8. Jump error monitoring

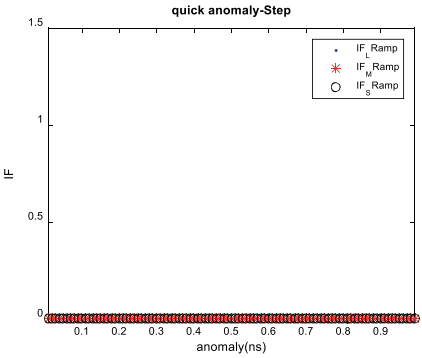


Fig. 9. Step error monitoring

The three-level slow-anomaly monitoring did not generate any flags, thus demonstrating that the monitoring algorithm is not affected by fast anomaly (i.e., the separation of faults with fast anomaly and slowly anomaly is realized).

5 Conclusions

Navigation satellite clock slow anomaly is one of the main parts of the in-orbit failure. In this paper, we study the slow anomaly of satellite clocks that have not been effectively monitored in orbit. We propose the satellite clock slow-anomaly SAIM method based on least-squares quadratic polynomial fitting to construct monitoring quantities. Fault monitoring based on the designed algorithm logic can be implemented on a single navigation satellite, and the TTA is short. Due to the separation of slow and fast anomaly, slow-anomaly monitoring will not be subject to fast anomaly, which will have a certain reference value for practical engineering applications.

Acknowledgments. The authors would like to give thanks to many other people in the China Academy of Space Technology (Xi'an) Satellite Navigation Institute for their advice.

References

1. Wang, Q.H., Droz, F., Rochat, P.: SpT on board atomic clocks and timing systems in GNSS. Geomat. Inf. Sci. Wuhan Univ. **36**(10), 1177–1181 (2011)
2. GPS Risk Assessment Study: Final Report. The Johns Hopkins University Applied Physics Laboratory, VS-99-007, January 1999
3. GEAS: GNSS Evolutionary Architecture Study, Phase I–Panel Report (2008)
4. GEAS: GNSS Evolutionary Architecture Study, Phase II–Panel Report (2010)
5. Tang, S., Liu, Y., Li, X.H.: A study on onboard satellite atomic clock autonomous integrity monitoring. J. Astronaut. **34**(1), 39–45 (2013)
6. Rodriguez-Perez, I., Garcia-Serrano, C., Catalan, C.C., Garcia, A.M., Tavella, P., Galleani, L., Amarillo, F.: Inter-satellite links for satellite autonomous integrity monitoring. Adv. Space Res. **47**(2), 197–212 (2011)
7. Bian, L., Meng, Y.S., Wang, X.L.: A principle design for global integrity of COMPASS. In: China Satellite Navigation Conference (CSNC 2012).2012.1
8. Bian, L., Liu, W.S., Yan, T., et.al.: Satellite Integrity Autonomous Monitoring (SAIM) of BDS and onboard performance evaluation. In: China Satellite Navigation Conference(CSNC 2018).2018.1
9. Heng, L.: GPS signal-in-space anomalies in the last decade: data mining of 400,000,000 GPS navigation messages. In: Proceedings of the 23rd International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS 2010), Portland, OR, September 2010, pp. 3115–3122 (2010)
10. Weiss, M., et al.: On board signal integrity for GPS. In: 23rd International Meeting of the Satellite Division of the ION, 21–24 Sept. 2010, Portland, OR (2010)

Massive MIMO Relaying with Hardware Impairments and ZF Processing

Sinian Jin[1], Dianwu Yue[1(✉)], and Meng Wang[1,2]

[1] College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China
{jinsinian,dwyue}@dlmu.edu.cn
[2] Department of Electronic Engineering, Dalian Neusoft University of information, Dalian 116026, China
wangmeng@neusoft.edu.cn

Abstract. In this paper, we consider the half-duplex massive MIMO relaying with the effect of hardware impairment. Depending on the half-duplex relay in amplify-and-forward (AF) scheme, multiple sources can simultaneously communicate with multiple destinations. In the case of requiring channel state information (CSI), relaying can use zero forcing (ZF) method to process receiving and transmitting signals. The achievable rate expression of relaying system with hardware impairment can be obtained. Depending on the deduced rate expression, the power scaling property can be analyzed in different cases. As the power of the source, uplink pilot and relay's transmitter are scaled, the results corroborate that when the number of antennas configured on the relaying grow without bound, the achievable rate can reach a fixed value.

Keywords: Massive MIMO relaying · Hardware impairment · Zero forcing · Amplify and forward · Power scaling law

1 Introduction

Massive MIMO is the most developed technology in 5G communication system. Based on its properties of improving spectrum efficiency and enhancing energy efficiency, it has became a hot topic in academia [1,2]. Since the relaying technology can simultaneously send and receive signals, the relaying system is a widely used technology, which have the advantage of reducing transmit power and improving system rate performance in mobile communications. Given the advantages of massive MIMO and relaying technology, the authors have considered combining the two key technologies for application [3–5].

Since massive MIMO needs to configure the large-scale antennas, it tends to be equipped with cheap transmitting and receiving components. But the low-quality components can have the effect of hardware impairments, it is a problem that cannot be ignored for massive MIMO systems. Specifically, [5–7] study the effect of low-quality components for massive MIMO system and find that

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 742–749, 2019.
https://doi.org/10.1007/978-981-13-6264-4_87

the influence of hardware impairments are given as an additive Gaussian noise with zero mean and variance depends on the transmitting and receiving energy. Against the above background, when the relaying use zero forcing (ZF) and amplify-and-forward (AF) scheme to process transmitting and receiving signals, this paper investigates the performance of massive MIMO relaying system in different power scaling laws.

2 System Model

2.1 Channel Model

In practice, with the help of half-duplex relaying K, source J_m, $m \in \{1, \cdots, M\}$, can send signals to destination L_m, where the relay's transceiver configured M_r and M_t antennas to serve each of the source and destination configured as single antenna. The channel matrix from the sources to the receiving antenna of relaying can be written as $\mathbf{T}_{JK} \in \mathbb{C}^{M_r \times M}$ and $\mathbf{T}_{KL} \in \mathbb{C}^{M_t \times M}$ is the channel from the destinations to the relay's transmitter. Specifically, $\mathbf{T}_\diamond$, $\diamond \in \{JK, KL\}$ can represent $\mathbf{T}_\diamond = \mathbf{G}_\diamond \mathbf{L}_\diamond^{1/2}$, where $\mathbf{G}_\diamond$ is the small-scale fading channel whose each element has $\mathcal{CN}(0,1)$ distribution, and $\mathbf{L}_\diamond$ represent the large-scale fading channel of diagonal matrix whose the mth diagonal element is $\gamma_{\diamond,m}$ [1,2].

2.2 Channel Estimation

During the phase of requiring channel state information (CSI), the pilot sequences of length ι ($\iota \geq M$) symbols send from each source and destination to the receiving terminal and transmitting terminal of the relay. Because the CSI of $\mathbf{T}_\diamond, \diamond \in \{JK, KL\}$ need to be estimated at the relay station, the received matrices $\mathbf{Y}_\diamond$ is [8]

$$\mathbf{Y}_\diamond = \sqrt{P_P}\mathbf{T}_\diamond \mathbf{O}_\diamond + \mathbf{T}_\diamond \mathbf{M}_{\eta,\diamond} + \mathbf{M}_{\theta,\diamond} + \mathbf{M}_\diamond \tag{1}$$

where P_P represent the uplink pilot power. $\mathbf{M}_\diamond$ are noise matrix at the relay, in which each element are independent and identically distributed (i.i.d) $\mathcal{CN}(0,1)$. $\mathbf{O}_\diamond \in \mathbb{C}^{M \times \iota}$ are the discrete Fourier transform pilot matrix which has the characteristics of $\mathbf{O}_\diamond \mathbf{O}_\diamond^H = \iota \mathbf{I}$ [8]. The imperfect radio-frequency (RF) chains can cause the distortion noise. The distortion noise on transmitter and receiver can represent $\mathbf{M}_{\eta,\diamond}$ and $\mathbf{M}_{\theta,\diamond}$, respectively. In particular, there is a important characteristic that the effect of distortion noise is proportional to the transmitted and received signal energy, respectively [5–7]. Thus, each elements of $\mathbf{M}_{\eta,\diamond}$ and $\mathbf{M}_{\theta,\diamond}$ are i.i.d. $\mathcal{CN}(0, \eta_P P_P)$ and $\mathcal{CN}\left(0, \theta_P \left(P_P (1+\eta_P) \sum_{m=1}^{M} \gamma_{\diamond,m} + 1\right)\right)$ random variables, where η_P and θ_P represent the level of transmit and receive imperfection, respectively. So the LMMSE estimation of $\mathbf{T}_\diamond$ is given by [8]

$$\hat{\mathbf{T}}_\diamond = \frac{\mathbf{Y}_\diamond}{\sqrt{P_P}} \left(\mathbf{O}_\diamond^H \mathbf{L}_\diamond \mathbf{O}_\diamond + \left((\eta_P + \theta_P + \theta_P \eta_P) \sum_{m=1}^{M} \gamma_{\diamond,m} + \frac{1+\theta_P}{P_P} \right) \mathbf{I} \right)^{-1} \mathbf{O}_\diamond^H \mathbf{L}_\diamond \tag{2}$$

With $\mathbf{T}_\diamond = \hat{\mathbf{T}}_\diamond + \boldsymbol{\varepsilon}_\diamond$, when η_P is small, we assume that $\hat{\mathbf{T}}_\diamond$ and $\boldsymbol{\varepsilon}_\diamond$ are independent. The mth column vectors of $\hat{\mathbf{T}}_\diamond$ and $\boldsymbol{\varepsilon}_\diamond$ are expressed as $\hat{\mathbf{t}}_{\diamond,m} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2_{\hat{\mathbf{t}}_{\diamond,m}}\mathbf{I}\right)$ and $\boldsymbol{\varepsilon}_{\diamond,m} \sim \mathcal{CN}(0, \sigma^2_{\boldsymbol{\varepsilon}_{\diamond,m}}\mathbf{I})$, respectively, where [2,8]

$$\sigma^2_{\hat{\mathbf{t}}_{\diamond,m}} = \frac{\iota\gamma^2_{\diamond,m}}{(\eta_P + \theta_P + \theta_P\eta_P)\sum\limits_{j=1}^{M}\gamma_{\diamond,j} + \frac{1+\theta_P}{P_P} + \iota\gamma_{\diamond,m}} \tag{3}$$

$$\sigma^2_{\boldsymbol{\varepsilon}_{\diamond,m}} = \gamma_{\diamond,m} - \sigma^2_{\hat{\mathbf{t}}_{\diamond,m}} \tag{4}$$

2.3 Downlink Transmission

When each source simultaneously transmits the signals $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ to the receiving terminal of the relay, the relaying broadcasts its processed signals by employed AF scheme to destinations. Thus, the signals of the relay's receiver be expressed as

$$\mathbf{y}_K = \sqrt{P_J}\mathbf{T}_{JK}\mathbf{x} + \mathbf{T}_{JK}\mathbf{m}_1 + \mathbf{m}_2 + \mathbf{m}_K \tag{5}$$

With AF scheme, the transmitted signal vector of relay is $\mathbf{s} = \sqrt{P_K}\rho\mathbf{F}\mathbf{y}_K$, where $\mathbf{F} = \hat{\mathbf{T}}_{KL}\left(\hat{\mathbf{T}}^H_{KL}\hat{\mathbf{T}}_{KL}\right)^{-1}\left(\hat{\mathbf{T}}^H_{JK}\hat{\mathbf{T}}_{JK}\right)^{-1}\hat{\mathbf{T}}^H_{JK}$ is the ZF processing matrix in AF scheme and ρ represents the power constraint factor. Therefore, the received signal of L_m is

$$y_{L,m} = \mathbf{t}^H_{KL,m}\mathbf{s} + \mathbf{t}^H_{KL,m}\mathbf{m}_3 + m_{L,m} + m_{4,m} \tag{6}$$

where $\mathbf{t}_{KL,m}$ is the mth column of $\mathbf{T}_{KL}$. The powers of source and the powers of transmitting terminal at the relay represent P_J and P_K, respectively. The noise vectors of receiving terminal at the relay and destinations can be represented as $\mathbf{m}_K \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, and $\mathbf{m}_L \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, respectively. The distortion noise vectors of sources, the receiving terminal at the relay, the transmitting terminal at the relay and destinations represent $\mathbf{m}_1$, $\mathbf{m}_2$, $\mathbf{m}_3$, and $\mathbf{m}_4$, respectively. The quantities x_m, $y_{L,m}$, $m_{1,m}$, $m_{4,m}$ and $m_{L,m}$ are the mth elements of $\mathbf{x}$, $\mathbf{y}_L$, $\mathbf{m}_1$, $\mathbf{m}_4$, and $\mathbf{m}_L$, respectively. Similarly, $y_{K,n}$, s_n, $m_{2,n}$, and $m_{3,n}$ are the nth elements of $\mathbf{y}_K$, $\mathbf{s}$, $\mathbf{m}_2$ and $\mathbf{m}_3$, respectively. Because the imperfect RF chain can have the effect of hardware impairment, the distortion noises be given by $m_{1,m} \sim \mathcal{CN}(0, \sigma_{1,m})$, $m_{2,n} \sim \mathcal{CN}(0, \sigma_{2,n})$, $m_{3,n} \sim \mathcal{CN}(0, \sigma_{3,n})$ and $m_{4,m} \sim \mathcal{CN}(0, \sigma_{4,m})$ with $\sigma_{1,m} = \eta_1 P_J \mathbb{E}\left\{|x_m|^2\right\}$, $\sigma_{2,n} = \theta_1 \mathbb{E}\left\{|y_{K,n}|^2\right\}$, $\sigma_{3,n} = \eta_2 \mathbb{E}\left\{|s_n|^2\right\}$ and $\sigma_{4,m} = \theta_2 \mathbb{E}\left\{|y_{L,m}|^2\right\}$, respectively, where η_1, η_2, θ_1 and θ_2 represent the levels of sources, the relay's transmitter, the relay's receiver and destinations imperfection, respectively. In this case, to meet the power constraint, $\mathbb{E}\left\{\|\mathbf{s}\|^2\right\} = P_K$, at the relay, the value of ρ is given by

$$\rho = \sqrt{\frac{1}{P_J\mathbb{E}\left\{\|\mathbf{F}\mathbf{T}_{JK}\|^2\right\} + \mathbb{E}\left\{\|\mathbf{F}\mathbf{T}_{JK}\mathbf{m}_1\|^2\right\} + \mathbb{E}\left\{\|\mathbf{F}\mathbf{m}_K\|^2\right\} + \mathbb{E}\left\{\|\mathbf{F}\mathbf{m}_2\|^2\right\}}} \tag{7}$$

3 Achievable Rate Analysis

3.1 Achievable Rate

By rewriting the received signal at L_m to the form of desired signal and noise, the achievable rate of lower bounded is as follows [3]:

$$\mathcal{S} \geq \frac{T-\iota}{T} \sum_{m=1}^{M} \log_2 \left(1 + \mathrm{S}INR_m\right) \tag{8}$$

where T represents the coherence interval. Thus, the SINR of L_m can be given by

$$\mathrm{S}INR_m = \frac{\rho^2 P_J P_K |W_{1,m}|^2}{\rho^2 P_J P_K (W_{2,m} + W_{3,m}) + \rho^2 P_K (W_{4,m} + W_{5,m} + W_{6,m}) + W_{7,m} + \sigma_{4,m} + 1} \tag{9}$$

where $W_{1,m} = \mathbb{E}\{\mathbf{t}_{KL,m}^H \mathbf{F} \mathbf{t}_{JK,m}\}$, $W_{2,m} = \mathbb{V}\mathrm{ar}\left(\mathbf{t}_{KL,m}^H \mathbf{F} \mathbf{t}_{JK,m}\right)$, $W_{3,m} = \sum_{j \neq m}^{M} \mathbb{E}\{|\mathbf{t}_{KL,m}^H \mathbf{F} \mathbf{t}_{JK,j}|^2\}$, $W_{4,m} = \mathbb{E}\left\{\left|\mathbf{t}_{KL,m}^H \mathbf{F} \mathbf{T}_{JK} \mathbf{m}_1\right|^2\right\}$, $W_{5,m} = \mathbb{E}\left\{\left|\mathbf{t}_{KL,m}^H \mathbf{F} \mathbf{m}_K\right|^2\right\}$, $W_{6,m} = \mathbb{E}\left\{\left|\mathbf{t}_{KL,m}^H \mathbf{F} \mathbf{m}_2\right|^2\right\}$ and $W_{7,m} = \mathbb{E}\left\{\left|\mathbf{t}_{KL,m}^H \mathbf{m}_3\right|^2\right\}$. In order to obtain $\mathrm{S}INR_m$, we define [2]

$$E_1 = \mathbb{E}\left\{\|\mathbf{F}\|^2\right\} \approx \sum_{m=1}^{M} \frac{1}{(M_t - M)(M_r - M)\sigma_{\hat{\mathbf{t}}_{KL,m}}^2 \sigma_{\hat{\mathbf{t}}_{JK,m}}^2} \tag{10}$$

$$E_{2,m} = \mathbb{E}\left\{\left\|\mathbf{t}_{KL,m}^H \mathbf{F}\right\|^2\right\} = \frac{1}{(M_r - M)\sigma_{\hat{\mathbf{t}}_{JK,m}}^2} + \sigma_{\varepsilon_{KL,m}}^2 E_1 \tag{11}$$

$$E_{3,m} = \mathbb{E}\left\{\|\mathbf{F}\mathbf{t}_{JK,m}\|^2\right\} = \frac{1}{(M_t - M)\sigma_{\hat{\mathbf{t}}_{KL,m}}^2} + \sigma_{\varepsilon_{JK,m}}^2 E_1 \tag{12}$$

$$E_{4,m} = \mathbb{E}\left\{\mathbf{t}_{KL,m}^H \mathbf{F} \mathbf{t}_{JK,m}\right\} = 1 \tag{13}$$

$$\begin{aligned} E_{5,mj} &= \mathbb{E}\left\{\left|\mathbf{t}_{KL,m}^H \mathbf{F} \mathbf{t}_{JK,j}\right|^2\right\} \\ &= \frac{\sigma_{\varepsilon_{JK,j}}^2}{(M_r - M)\sigma_{\hat{\mathbf{t}}_{JK,m}}^2} + \frac{\sigma_{\varepsilon_{KL,m}}^2}{(M_t - M)\sigma_{\hat{\mathbf{t}}_{KL,j}}^2} + \sigma_{\varepsilon_{KL,m}}^2 \sigma_{\varepsilon_{JK,j}}^2 E_1 + \sigma_{mj} \end{aligned} \tag{14}$$

where σ_{mj} is the Dirac function (i.e., $\sigma_{mj} = 1$ if $j = m$ and 0 otherwise). Through the above, we also $\sigma_{1,m} = \eta_1 P_J$, $\sigma_{2,n} = \theta_1 \left(P_J(1+\eta_1) \sum_{m=1}^{M} \gamma_{JK,m} + 1\right)$, $\sigma_{3,n} = \frac{\eta_2 P_K}{M_t}$ and

$$\sigma_{4,m} = \theta_2 \left(\rho^2 P_K P_J (1+\eta_1) \sum_{j=1}^{M} E_{5,mj} + \rho^2 P_K (1 + \sigma_{2,n}) E_{2,m} + \eta_2 P_K \gamma_{KL,m} + 1\right) \tag{15}$$

$$\rho = \sqrt{\frac{1}{(1+\sigma_{2,n})\, E_1 + P_J\,(1+\eta_1) \sum\limits_{m=1}^{M} E_{3,m}}} \tag{16}$$

We can substitute $W_{1,m} = E_{4,m}$, $W_{2,m} = E_{5,mm} - E_{4,m}^2$, $W_{3,m} = \sum\limits_{j\neq m}^{M} E_{5,mj}$, $W_{4,m} = \sigma_{1,m} \sum\limits_{j=1}^{M} E_{5,mj}$, $W_{5,m} = E_{2,m}$, $W_{6,m} = \sigma_{2,n} E_{2,m}$, $W_{7,m} = \eta_2 P_K \gamma_{KL,m}$, $\sigma_{4,m}$ and ρ into (9) and obtain the expression of achievable rate.

3.2 Power Scaling Laws

Based on (8), we give the power scaling laws in two case. To better understand the system characteristics in the power scaling, we define the limiting case of $M_r \to \infty$, $M_t \to \infty$, and $M_r = \kappa M_t$, $\kappa > 0$. In this limiting case, we make $P_J = \overline{P}_J / M_r^a$, $P_K = \overline{P}_K / M_t^a$ and $P_P = \overline{P}_P / M_r^b$, where $\overline{P}_J$, $\overline{P}_K$ and $\overline{P}_P$ are fixed.

Corollary 1: When $a = 1$ and $b = 0$, the rate expression of (8) is given by

$$S^{case1} \to \frac{T-\iota}{T} \sum_{m=1}^{M} \log_2 \left(1 + \frac{\bar{P}_J \bar{P}_K}{\bar{P}_J \bar{P}_K \Delta_1 + \bar{P}_J \Delta_2 + \bar{P}_K \Delta_3 + \Delta_4}\right) \tag{17}$$

where $\Delta_1 = \eta_1 + \theta_2 + \eta_1\theta_2$, $\Delta_2 = (1+\eta_1)(1+\theta_2) \sum\limits_{j=1}^{M} \sigma^{-2}_{\hat{\mathbf{t}}_{KL,j}}$, $\Delta_3 = (1+\theta_1)(1+\theta_2)\, \sigma^{-2}_{\hat{\mathbf{t}}_{JK,m}}$ and $\Delta_4 = (1+\theta_1)(1+\theta_2) \sum\limits_{j=1}^{M} \sigma^{-2}_{\hat{\mathbf{t}}_{KL,j}} \sigma^{-2}_{\hat{\mathbf{t}}_{JK,j}}$.

Corollary 2: When $a + b = 1$ and $0 < a < 1$, the rate expression of (8) can be expressed as

$$S^{case2} \to \frac{T-\iota}{T} \sum_{m=1}^{M} \log_2 \left(1 + \frac{\bar{P}_J \bar{P}_K}{\bar{P}_J \bar{P}_K \Delta_1 + \bar{P}_J \Delta_5 + \bar{P}_K \Delta_6 + \Delta_7}\right) \tag{18}$$

where $\Delta_5 = (1+\eta_1)(1+\theta_2) \sum\limits_{j=1}^{M} \frac{(1+\theta_P)\kappa^b}{\iota \bar{P}_P \gamma^2_{KL,j}}$, $\Delta_6 = \frac{(1+\theta_1)(1+\theta_2)(1+\theta_P)}{\iota \bar{P}_P \gamma^2_{JK,m}}$ and $\Delta_7 = (1+\theta_1)(1+\theta_2) \sum\limits_{j=1}^{M} \frac{(1+\theta_P)^2 \kappa^b}{\iota^2 \bar{P}_P^2 \gamma^2_{KL,j} \gamma^2_{JK,j}}$. By (17) and (18), we can find $S^{case2} > S^{case1}$.

4 Simulation Results

For comparison with (8), the exact achievable rate can be given by [3]

$$\tilde{\mathcal{S}} \triangleq \frac{T-\iota}{T} \sum_{m=1}^{M} \mathbb{E}\left\{\log_2\left(1 + \widetilde{SINR}_m\right)\right\} \tag{19}$$

where

$$\widetilde{SINR}_m = \frac{\rho^2 P_J P_K \left|\mathbf{t}_{KL,m}^H \mathbf{F}\mathbf{t}_{JK,m}\right|^2}{\left(\begin{array}{l}\rho^2 P_J P_K \sum\limits_{j\neq m}^{M} \left|\mathbf{t}_{KL,m}^H \mathbf{F}\mathbf{t}_{JK,j}\right|^2 + \rho^2 P_K \left|\mathbf{t}_{KL,m}^H \mathbf{F}\mathbf{T}_{JK}\mathbf{m}_1\right|^2 + \sigma_{4,m} \\ +\rho^2 P_K \left|\mathbf{t}_{KL,m}^H \mathbf{F}\mathbf{m}_K\right|^2 + \rho^2 P_K \left|\mathbf{t}_{KL,m}^H \mathbf{F}\mathbf{m}_2\right|^2 + \left|\mathbf{t}_{KL,m}^H \mathbf{m}_3\right|^2 + 1\end{array}\right)} \tag{20}$$

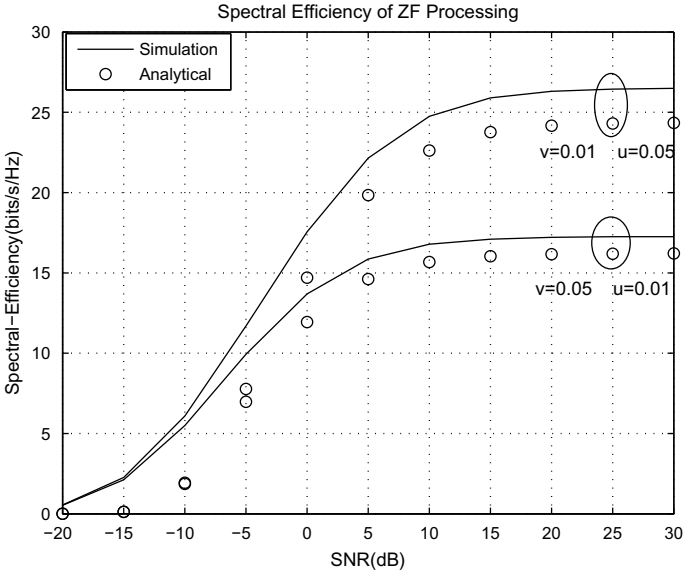


Fig. 1. The achievable rate versus SNR.

Throughout the examples, we assumed that $T = 200$, $M = 5$, $\iota = M$, $\mathbf{L}_{JK} = \mathbf{L}_{KL} = \mathbf{I}$ and $M_r = M_t$. Furthermore, we define the level of imperfect hardware quantify at sources and destinations are $\eta_1 = \eta_P = \theta_2 = v$ and the level of imperfect hardware quantify at the receiving and transmitting terminal of the relay are $\eta_2 = \theta_P = \theta_1 = u$, respectively. To better examine the system performance of using inexpensive hardware, the value of imperfect hardware quantify are chosen to be higher than 0.175^2.

First, Fig. 1 compares the achievable rate for the analytical results based on (8) and numerical results based on (19). Here, Fig. 1 shows the achievable rate versus signal-to-noise ratio (SNR) in the case of different levels of hardware impairments. We set $M_r = M_t = 100$ and $P_J = P_K = P_P = SNR$ in this example. We can clearly see that the analytical results in (8) close to the simulation results in (19) and it suggests that the derived theoretical expression in (8) is accurate. By comparing with the situation of using high-quality components of the relay, we also find that using high-quality components of sources and destinations can effectively increase the rate system.

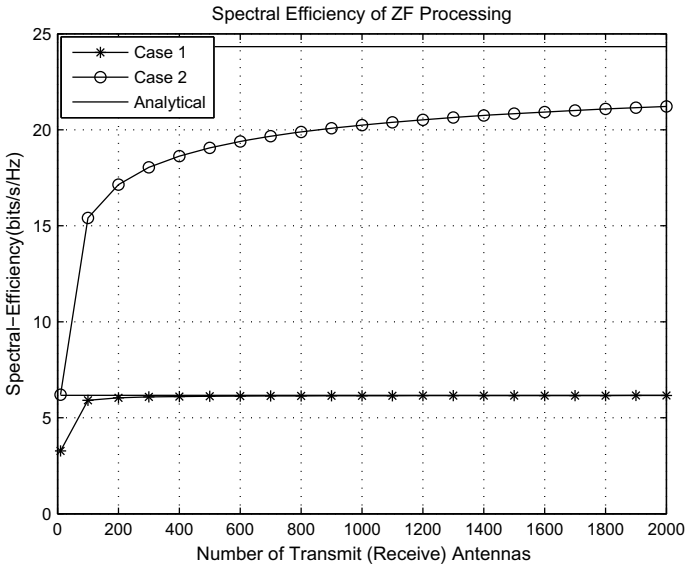


Fig. 2. The achievable rate versus the number of relay's antennas.

In the case of $\bar{P}_J = \bar{P}_K = \bar{P}_P = 10\,\mathrm{d}B$ and $v = u = 0.01$, we choose two typical power scaling cases: Case 1: $P_J = \frac{\bar{P}_J}{M_r}$, $P_K = \frac{\bar{P}_K}{M_t}$ and $P_P = \bar{P}_P$; Case 2: $P_J = \frac{\bar{P}_J}{\sqrt{M_r}}$, $P_K = \frac{\bar{P}_K}{\sqrt{M_t}}$ and $P_P = \frac{\bar{P}_P}{\sqrt{M_r}}$. The curve of achievable rate in two power scaling cases are plotted in Fig. 2. It is clear that as M_r and M_t grow large, the two achievable rate curves approach to (17) and (18), which are given by the corresponding constant asymptotical value, and show that our asymptotical results are accurate. Finally, we also can find that the asymptotical value in Case 1 is less than that in Case 2. Such relative performance comparison agrees with the corollaries.

5 Conclusion

This paper investigated the massive MIMO relaying with hardware impairments, while the relaying use ZF method and AF scheme to process signal. When considering the effect of imperfect CSI, the paper derived the achievable rate of relaying with hardware impairment and deduced the rate properties in different cases of power scaling. Simulation results shown the accuracy of the derived expression and also illustrated that comparing with the high-quality components of the relay, the high-quality components of sources and destinations have a significant boost for system performance.

References

1. Marzetta, T.L.: Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans. Wirel. Commun. **9**(11), 3590–3600 (2010)
2. Ngo, H.Q., Larsson, E.G., Marzetta, T.L.: Energy and spectral efficiency of very large multiuser MIMO systems. IEEE Trans. Commun. **61**(4), 1436–1449 (2013)
3. Ngo, H.Q., Larsson, E.G.: Large-scale multipair two-way relay networks with distributed AF beamforming. IEEE Commun. Lett. **17**(12), 1–4 (2013)
4. Ngo, H.Q., Suraweera, H.A., Matthaiou, M., et al.: Multipair full-duplex relaying with massive arrays and linear processing. IEEE J. Sel. Areas Commun. **32**(9), 1721–1737 (2014)
5. Jin, S.-N., Yue, D.-W., Nguyen, H.H.: Power scaling laws of massive MIMO full-duplex relaying with hardware impairments. IEEE Access **6**, 40860–40882 (2018)
6. Björnson, E., Hoydis, J., Kountouris, M., et al.: Massive MIMO systems with non-ideal hardware: energy efficiency, estimation, and capacity limits. IEEE Trans. Inf. Theory **60**(11), 7112–7139 (2014)
7. Björnson, E., Matthaiou, M., Debbah, M.: Massive MIMO with non-ideal arbitrary arrays: hardware scaling laws and circuit-aware design. IEEE Trans. Wirel. Commun. **14**(8), 4353–4368 (2015)
8. Biguesh, M., Gershman, A.B.: Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals. IEEE Trans. Signal Process. **54**(3), 884–893 (2006)

Performance Comparisons on OFDM System with Different Channels

Qin Liu(✉), Xueguang Yuan, Yangan Zhang, Yuhua Zhang, and Yalei Chu

State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China
18801073071@163.com, {yuanxg, zhang, cyl}@bupt.edu.cn, zyh265@qq.com

Abstract. We consider the impact of cyclic prefix (CP) length to the BER under the Gaussian channel and the Rayleigh channel with different SNRs on orthogonal frequency-division multiplexing (OFDM) system, respectively. The BER gap between the two channels can reach 5 ~ 10 times higher at the equal CP length.

OCIS Codes: (060.4510) optical communication · (220.4830) systems design

1 Introduction

Orthogonal frequency-division multiplexing (OFDM) is extensively applied to optical communication for the flexibility of spectrum assignment and the high spectral efficiency, as mentioned in paper [1, 2]. The BER of the OFDM system is impacted by subcarrier number, cyclic prefix (CP) length, channel conditions, and so on. Nevertheless, in paper [3], changing the number of subcarriers will increase computational complexity and only get about 1 dB gain in Eb/N0. Besides, a symbol with CP can restrain the intersymbol interference (ISI) and inter-carrier interference (ICI) from dispersive channel [4]. Therefore, discussing the impact of CP length to the BER under different channels is necessary [5].

In this paper, due to the strong anti-noise of Quadrature Amplitude Modulation (QAM), we compare the BER under the Gaussian channel and the Rayleigh channel on 16QAM-OFDM systems by changing the CP length, and our results achieve that, for the same CP length, the BER under the Gaussian channel is nearly 5 ~ 10 times lower than that of the Rayleigh channel. In particular, when SNR is greater than 14 dB, the BER gap can be at least 10 times higher.

2 OFDM Simulation System Model and Theory

As mentioned above, we change the CP length to compare the BER between the two different channels based on MATLAB. A general overview of 16QAM-OFDM simulation model is shown in Fig. 1.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 750–753, 2019.
https://doi.org/10.1007/978-981-13-6264-4_88

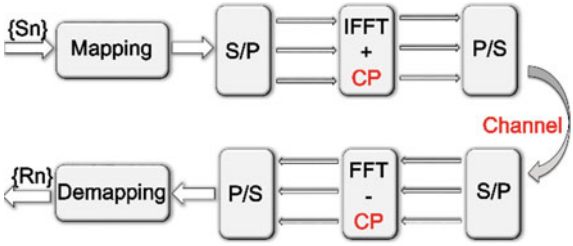


Fig. 1. Simulation model of 16QAM-OFDM system

In order to reduce the interference, CP samples copied from the end of a symbol are placed in the front of the symbol. The specific schematic diagram is Fig. 2. The duration of an OFDM symbol is T_{IFFT} and the CP is T_{cp}. Because of the long CP, channel convolution is equivalent to the cyclic convolution, and subcarriers are still overlapped and orthogonal over the duration of T. For an OFDM system, the carriers have spectral nulls at all the other carrier frequencies. These spectral nulls are illustrated in Eq. (1). Therefore, OFDM systems are able to maximize spectral efficiency without causing adjacent channel interference.

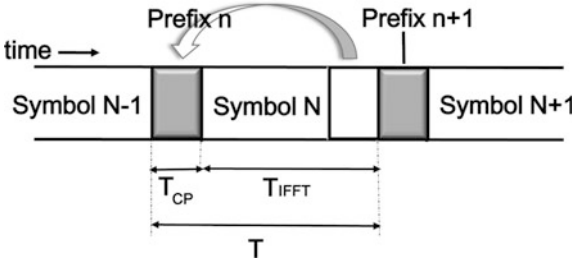


Fig. 2. Addition of cyclic prefix to an OFDM symbol

$$\int_0^{T_{FFT}} \cos(2\pi mt/T_{FFT})\cos(2\pi nt/T_{FFT})dt = \begin{cases} \frac{T_{FFT}}{2} & m = n \\ 0 & m \neq n \end{cases} \tag{1}$$

$$y(n) = x(n) * h(n) = \sum_{m=0}^{L_h-1} x(m) * h(n-m) \tag{2}$$

At the same time, when the transmission symbols $x(n)$ with the length of $N + L$ pass through the channel with impulse response length of L_h, the receiving signal is shown in Eq. (2). A part of $y(n)$ comes from the interference of the previous symbol.

3 Simulation Results

We simulated the 16QAM-OFDM system with 512 subcarriers under the Gaussian channel and the Rayleigh channel. Then, comparing the BER curves with different CP lengths in small range and large range, respectively. As Fig. 3 (a) and (b) shows, when CP length ranges from 0T to 0.04T, the BER decreases with the increase of SNR. For the same CP length, the BER under the Gaussian channel is nearly 5 ~ 10 times lower than that of the Rayleigh channel. Especially, when SNR is greater than 14 dB, the BER gap can reach at least 10 times higher. Nevertheless, in Fig. 3 (c), the BER curves are overlapped in large scale. From Fig. 3 (d), we can know that BER is lower than 0.1 when SNR is greater than 16 dB under the Rayleigh channel. Therefore, the BER under the Gaussian channel is far better than another.

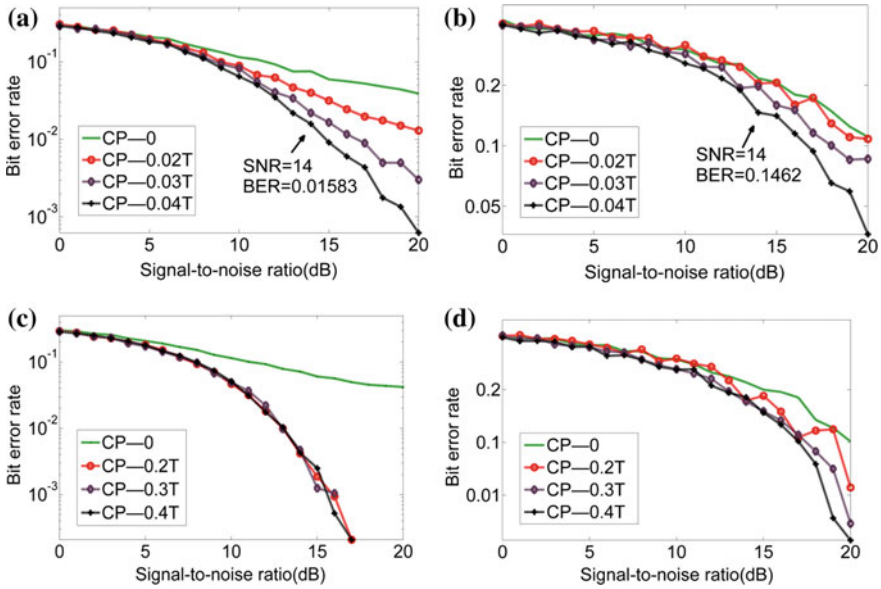


Fig. 3. (a) BER versus SNR under Gaussian channel with different CP lengths in the small range. (b) BER versus SNR under Rayleigh channel with different CP lengths in small range. (c) BER versus SNR under Gaussian channel with different CP length in the large range. (d) BER versus SNR under Rayleigh channel with different CP length in the large range

In Fig. 4, the performance of 16QAM-OFDM system is illustrated in different CP guard interval and signal-to-noise ratio. With increasing SNR and CP length, the BER of the system is improved as illustrated.

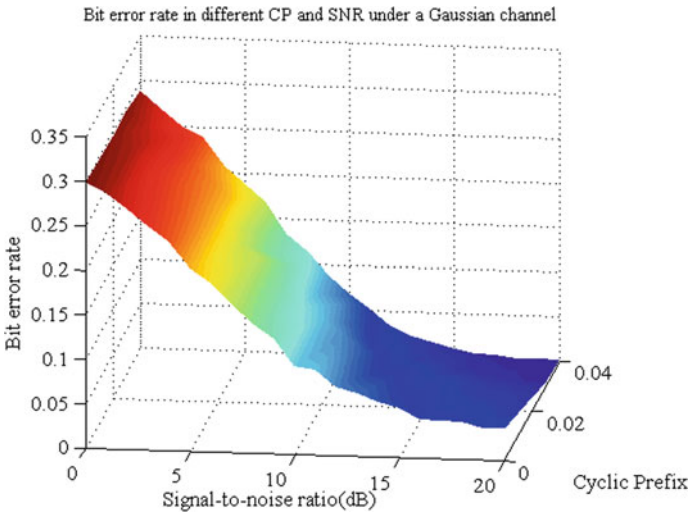


Fig. 4. Bit error rate in different CPs and SNRs under a Gaussian channel

4 Conclusion

The simulation of 16QAM-OFDM system based on MATLAB provides a better way to comprehend the impact of CP length to the BER under the two different channels. In the end, the BER under the Gaussian channel can reach $5 \sim 10$ times lower than that of the Rayleigh channel at the equal CP length.

References

1. Hu, R., et al.: Digital OFDM-PON employing binary intensity modulation and direct detection channels. In: 2017 Optical Fiber Communications Conference and Exhibition (OFC), Los Angeles, CA, pp. 1–3 (2017)
2. Shao, Y., et al.: A novel subcarrier OFDM-MSK WDM passive optical network. In: 2010 OFC, San Diego, CA, pp. 1–3 (2010)
3. Baltar, L.G., Nossek, J.A.: Multicarrier systems: a comparison between Filter Bank based and Cyclic Prefix based OFDM, OFDM 2012. In: Proceedings of 17th International OFDM Workshop 2012 (InOWo'12), Essen, Germany, pp. 1–5 (2012)
4. Mohammed, U.S., Tohamy, A.: A low complexity OFDM system with minimum intersymbol interference. In: 2012 5th Joint IFIP Wireless and Mobile Networking Conference (WMNC), Bratislava, pp. 22–29 (2012)
5. Nourollahi, H., Maghrebi, S.G.: Evaluation of cyclic prefix length in OFDM system based for Rayleigh fading channels under different modulation schemes. In: 2017 ISCC, Heraklion, pp. 164–169 (2017)

The Influence of Obstacle Blockage on 5G Millimeter Wave Communication

Linlin Xu[1,2] and Dianwu Yue[1(✉)]

[1] Dalian Maritime University, Dalian, China
dwyue@hotmail.com
[2] University of Science and Technology Liaoning, Anshan, China

Abstract. Millimeter wave (mmWave) communication is one of the hottest topics in the fifth- generation communication (5G) research. The influence of blockages on mmWave communication is very serious. This paper provides an overview of the research status of the blockage effect. First, it briefly introduces the impact of natural environment blockages and human blockages. Second, it focuses on the building blockage effect which is the most serious blocking form of cellular networks communication, and then discusses a tractable stochastic analysis approach for modeling the building blockage effect. Combining with the system performance analysis, some more practically simplified methods for the complex system models are condensed. Finally, the direction that can be studied to solve obstacle blockage effect is proposed.

Keywords: Blockage effect · Millimeter wave · Stochastic geometric theory · Cellular networks

1 Introduction

With the unprecedented development of the Internet, the demand for wireless network resources increases explosively, the microwave communication becomes serious congestion, and the millimeter wave communication with wider frequency band becomes the research hotspot. However, due to the presence of obstacles in the environments, the blocking effect seriously affects the transmission quality of wireless communication, especially for the signal in the mmWave frequency band. For example, the penetration loss of mmWave through human body is 20–40 dB [1], and the loss of mmWave through buildings is as high as 40–80 dB [2]. The mmWave is more sensitive to the blockage effect. The longer the signal transmission path is, the more intersecting obstacles may be and the more obvious the attenuation is. It will bring great error to the analysis of mmWave performance. But the blockage effect in the traditional system model is often ignored in the analysis, or incorporated into the shadow model with reflection, scattering, and diffraction.

The commonly used cellular network model is the hexagonal grid model. The base station(BS) is fixed and located at the center of the hexagonal grid.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 754–761, 2019.
https://doi.org/10.1007/978-981-13-6264-4_89

However, this model is too idealized to be applied to irregular landforms, such as mountains, lakes, canyons, and so on. At the same time, with the construction of urban network, in the face of a variety of network coverage needs, such as shopping malls, airports and other crowded hot areas, the location of macro BS and micro BS needs to have a lot of random. In addition, the deployment of low-power nodes based on various requirements reduces the size of the cellular network [3] and rarely adapts to the regular grid topology. Therefore, a stochastic geometric model of a randomly distributed BS is closer to the characteristics of a real cellular networks [4–9].

In the second and third parts, this paper briefly introduces the blockage effect of natural environments and human body to mmWave communication. The fourth part takes the blockage effect of buildings as an example, the description method and analysis method of stochastic geometry theory in analyzing obstacle blockage effect are described in detail. Last, we discuss the open research challenges and future research directions.

2 Blockage Effect of Natural Environments

Natural environments, such as atmospheric particles, rainwater, fog, cloud layer, temperature and leaves, and other factors all have different effects on mmWave communication [10], different geographical features such as rainy areas, there are great differences in the jungle area. The following three typical blocking effects of atmospheric attenuation, rainwater attenuation, and vegetation attenuation are introduced, respectively.

Atmospheric attenuation: Atmospheric particles absorb and scatter mmWave signals, which can cause the attenuation of signals exponentially with the distance between the transmitter and the receiver. However, with the exception of the worst environments, such as tropical rainforests, the atmospheric particles have very little attenuation effect on mmWave because the expected cell size of the 5G system is less than 200 m [11].

Rain attenuation: Compared with long wavelength signal, rainwater can attenuate short wavelength signal, so the attenuation of rainwater to mmWave transmission is more serious in all natural environmental blockages that affect 5G communication. The degree of attenuation caused by rain depends on rainfall. Small rain can cause accidental errors in the received data. Rainstorm can even cause the interruption of mmWave signal transmission [10]. In the current literature, for the case of obvious rain, the wobulation technique can make the received data error smaller, but the blockage effect caused by rainstorm has not been further explored [10].

Vegetation attenuation: mmWave communication is extremely sensitive to the blockage of leaves. In the measurement of 73 GHz broadband mmWave, the attenuation of the outdoor leaves to mmWave signals is 0.4 dB/m [12], and attenuation usually increases with the increase of the signal path through the leaf path [13]. The dependence of attenuation coefficient on tree type, season, and BS elevation is a continuous research topic [14].

3 Blockage Effect of Human Body

Most mmWave communications involve the transmission of information with mobile users, so the human body itself is an unavoidable severity blockage, and the higher the signal frequency, the greater the attenuation [17]. At the same time, some people active near the mobile terminal can also have an intermittent blockage effect on communication, which will change the signal on a very small time scale.

According to the characteristics of human body, the cylinder [15,16], multiple edge [17], and other models are often used to represent the human body structure. The single person blockage, multi-person blockages and many people blocking each other are studied. The theoretical results are basically consistent with the experimental data. When the human body moves horizontally between the transmitter and the receiver, the midpoint of the line attenuates the least, and when more than one person passes through the transmitter and the receiving terminal, the attenuation of the transverse passage of the human body is greater than that of the lateral passage of the human body [17].

At present, wearable products, such as smart glasses, are popular in the research of mmWave communication. The blockage effect of human body on wearable products is closely related to the location of users, such as whether it is in the corner or in the direction of the body. The use of directional antennas can improve the system gain and minimize interference. The position of the wearable product may be at any height [18], so the blocking model applicable to any location needs to be further studied.

4 Blockage Effect of Buildings

At present, the social environments is constantly changing, there will be a variety of changing buildings around for a period of time, bustling urban center, high-rise buildings are abound. However, the random appearance of buildings, especially the buildings with a relatively high building, has a very serious impact on mmWave communication. The measurement results show that the loss of mmWave penetrating buildings is as high as 40–80 dB [2], which is usually difficult to overcome. Therefore, indoor and outdoor mmWave communication networks are independent of each other and need to be discussed separately [19]. At the same time, it is often assumed that the signal cannot penetrate the building, that is, the signal is completely attenuated after passing through the building, so that when there is no building blocking between the BS and the user, the mmWave communication path is considered as the line-of-sight (LOS) communication path. On the contrary, it is a non-line-of-sight (NLOS) communication path. The signal in the LOS path transmission more like transmission in free space, the path loss exponent is about 2 [1], the signal receiver receives up to 60%–80% from the LOS path [16], NLOS is mainly used in the signal beam reflection on the surface of the building, behind the overlay network blocking shadow area [1], The NLOS path loss exponent and environmental factors are

closely related, and is always greater than the LOS path loss exponent, such as the NLOS path loss exponent in New York City 5.76 [20]. Therefore, compared to an NLOS BS, the user is likely to connect to a BS with a farther LOS path and the communication quality is higher.

In the outdoor mmWave communication situation, the traditional cellular network model is used in the hexagonal lattice structure, cell location, and the BS are relatively fixed in position, but this model is too idealistic, not suitable for irregular topography, and has great random of the macro and micro characteristics of BS, therefore, the random geometric model of a random distributed BS is more close to the characteristic of the actual cellular networks [4–9].

4.1 Building Block Model Based on Stochastic Geometry Theory

In the two-dimensional space, the performance analysis of the system is based on the typical user as the reference point, as the green triangle shown in Fig. 1 is a typical user, the location of the BS is the homogeneous PPP (Poisson Point Process) random distribution, the violet circle is used to represent the LOS BS and the NLOS BS is blocked by the building, the serving BS of the user is the nearest LOS BS that is the closest to the user. Thus, the Voronoi cell model with random polygon boundary is formed, the location of the BS is random. Every user is communicating with the nearest LOS BS. So, the size and shape of the cell is also random, and the edge effect of the cell is reduced [8].

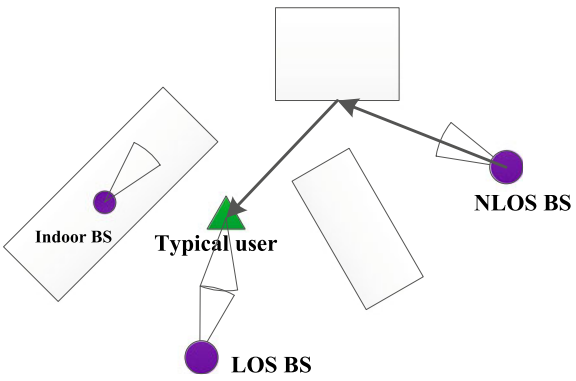


Fig. 1. Random buildings blockage model

The position of the block is characterized by the central position of the building, the location of the building is also the Poisson point process, and is independent of the distribution of the BS, the shape of a building is represented by a rectangle. The length and width of the rectangle are represented by the independent and identically distributed (i.i.d.) random variables. The height parameters of the building have little effect on the final result, only one scaling

factor, so the model does not join the height parameter. When the intersection between LOS path and blocking building forms Poisson random process, the scaling factor is a constant. It can be easily modified to take into account the height [5], so the height of buildings is ignored. The direction of the building is represented by the uniformly distributed random variable, and the model is the stochastic geometric blockage model [4–9].

4.2 System Performance Description

In the cellular network, the coverage of the system is used to describe the performance of the system, which is defined as the probability that the signal-to-interference-plus-noise ratio (SINR) is greater than a specific threshold. The coverage of the system can represent the range of cells in which the received SINR is larger than the threshold. It can also be understood as the probability of the number of random users who can realize the threshold of SINR. The coverage of the system is closely related to the density of the BS, the density of building blockages, the average length and width of the buildings, and so on, especially the connection strategy between users and BS. For example, users communicate with the nearest NLOS BS, and the path loss of the system is higher than the remote LOS BS, and the SINR of the system is low. So users have to connect to the nearest LOS BS. When the SINR of the nearest LOS BS downlink communication is greater than the threshold, the LOS BS can cover the user.

How to build a connection between a user and the nearest LOS BS is to be considered from three aspects, first, it is determined that the LOS BS is present, and the BS which is communicated with the user is the LOS BS, that is, the building blockage does not appear on the communication path. Secondly, the nearest BS is selected from the alternative LOS BS, and the other LOS BS is the largest interfering BS in the communication. Finally, the distance distribution between the nearest LOS BS and the user can be expressed by the LOS probability function, which is the exponential function of distance. As the distance becomes longer, the LOS path is more likely to be blocked by buildings. So the LOS probability function is a decreasing function. A connection between the user and the nearest LOS BS is established to derive the coverage of the system under this restricted condition.

The analysis of SINR coverage in the system, assuming that the fixed transmission power of each BS ,by means the array gain of large scale antenna array directional beamforming, is provided at transmitter and receiver compensation frequency correlation path loss to reduce interference, A Rayleigh fading or a Nakagami fading can be used to model the effect of the small-scale multipath caused by reflection and scattering. Path loss is modeled as an exponential function about distance. The system interference is mainly caused by other LOS BS and Additive White Gaussian Noise (AWGN).

4.3 Simplified Methods of Complex System Model

Simplification of the building block model: Because the length and direction of blocking buildings play an important role in the channel, the width of the rectangle can be neglected. The rectangular model is replaced by the line segment model [7]. The simulation results of coverage are the same, and the line segment model will be simpler, having a small effect on the results.

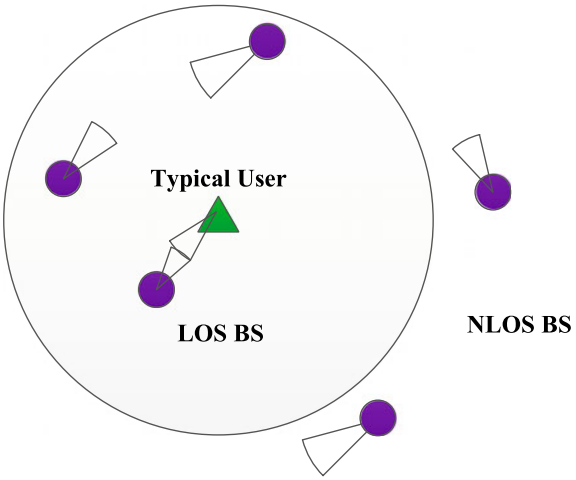


Fig. 2. Approximation using the equivalent LOS ball

The simplification of the LOS probability function: The LOS area of a typical user can be approximated by its equivalent LOS sphere, as shown in Fig. 2, by replacing the LOS probability function with a step function, usually with a radius of 200 m. When the distance between the LOS BS and the user is less than the radius of the LOS sphere, the LOS probability function is 1, and the other cases are zero. By using this simplified approximation method, the lower limit of the actual SINR distribution is usually obtained. When the BS density increases [9], the error caused by the approximation algorithm will become smaller, and at the same time, this approximation simplifies the expression of complex variable limit multiple integrals. The system coverage can be estimated more quickly and the system performance can be analyzed more easily.

Simplification of dense networks: If the number of LOS BS observed by typical users is larger than the number of building blockages, then the mmWave cellular network is a dense network. In the analysis, the influence of NLOS BS and thermal noise can be ignored, and the performance of the system is limited by the interference of other LOS BSs. Simulation shows that this negligible error is introduced in performance evaluation [9]. Because it is found that the power of the signal from the nearby mmWave LOS transmitter is almost deterministic

[20], the influence of small-scale fading can be ignored in the analysis of the dense network at the same time. Although there are some deviations at high SINR, the simplified analysis model is usually closer to the real environments [9].

5 Conclusion

This paper expounds the influence of obstacle blockages on mmWave wireless communication from the aspects of natural environment blockages, human body blockages, and building blockages. Some of these blockage effects are negligible, some blocking effects are sporadic and need further observation and research, and some blockage effects are persistent and seriously threaten the normal communication quality. In this paper, building blockages are taken as an example to describe in detail the blockage effects caused by obstacles to mmWave wireless communication. A tractable stochastic geometric method is introduced, which can be used in many ways to model blockages that is more in line with cellular network characteristics, and the method of analysis of system performance coverage under this model is discussed. At the same time, the simplification methods for complex system models under various conditions are introduced.

In future work, it would be interesting to analyze the correlation between blocking paths, as the influence of the same building blockage on different communication paths is not completely independent, for example, if two base stations and the same user happen to be blocked by the same building, the two paths are related to each other. It would be another interesting topic to incorporate the nonconvex building blockage effects. A direct path can intersect more than once with a nonconvex building, so the blockage effect is different from the convex building.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under Grants 61371091, 61301228.

References

1. Rajagopal, S., Abu-Surra, S., Malmirchegini, M.: Channel feasibility for outdoor non-line-of-sight mmWave mobile communication. In: 2012 IEEE Vehicular Technology Conference, pp. 1–6. Canada (2012)
2. 5G Channel Model for Bands Up to 100 GHz: Aalto University. Espo, Finland, White Paper (2015)
3. Ghosh, A. et al.: Heterogeneous cellular networks. From theory to practice. IEEE Commun. Mag. **50**(6), 54–64 (2012)
4. Bai, T., Robert Jr., W.H.: Coverage analysis for millimeter wave cellular networks with blockage effects. In: Global Conference on Signal and Information Processing, pp. 727–730. USA (2013)
5. Bai, T., Vaze, R., Robert Jr., W.H.: Analysis of blockage effects on urban cellular networks. IEEE Trans. Wirel. Commun. **13**(9), 5070–5083 (2014)
6. Akoum, S., Ayach, E.O., Robert Jr., W.H.: Coverage and capacity in mmwave cellular systems. In: 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers, pp. 688–692. USA (2012)

7. Bai, T., Vaze, R., Robert Jr., W.H.: Using random shape theory to model blockage in random cellular networks. In: 2012 International Conference on Signal Processing and Communications, pp. 1–5. India (2012)
8. Jeffrey, G.A., Baccelli, F., Krishna Ganti, R.: A tractable approach to coverage and rate in cellular networks. IEEE Trans. Commun. **59**(11), 3122–3134 (2011)
9. Bai, T., Robert Jr., W.H.: Coverage and rate analysis for millimeter wave cellular networks. IEEE Trans. Commun. **14**(2), 1100–1114 (2015)
10. Neha, P.N., Dipashree, M.B.: A survey on obstacles for 5G communication. In: 2015 International Conference on Communications and Signal Processing, pp. 0831–0835. India (2015)
11. Mansoor, S., et al.: 5G: a tutorial overview of standards, trials, challenges, deployment, and practice. IEEE J. Sel. Areas Commun. **35**(6), 1201–1221 (2017)
12. Theodore, S.R., Deng, S.: 73 GHz wideband millimeter-wave foliage and ground reflection measurements and models. In: 2015 IEEE International Conference on Communication Workshop, pp. 1238–1243. India (2015)
13. Schwering, F.K., Violette, E.J., Espeland, R.H.: Millimeterwave propagation in vegetation: experiments and theory. IEEE Trans. Geosci. Remote Sens. **26**(3), 355–367 (1988)
14. Haneda, K., et al.: 5G 3GPP-like channel models for outdoor urban microcellular and macrocellular environments. In: 2016 IEEE 83rd Vehicular Technology Conference, pp. 1–7. Nanjing (2016)
15. Vitaly, P., et al.: Interference and SINR in millimeter wave and terahertz communication systems with blocking and directional antennas. IEEE Trans. Wirel. Commun. **16**(3), 1791–1808 (2017)
16. Gapeyenko, M., et al.: Analysis of human body blockage in millimeterwave wireless communications systems. In: IEEE International Conference on Communications, pp. 1–7. Malaysia (2016)
17. Zhao, X., et al.: Attenuation by human bodies at 26- and 39.5-GHz millimeter wavebands. IEEE Antennas Wirel. Propag. Lett. **16**, 1229–1232 (2017)
18. Venugopal, K., Robert Jr., W.H.: Location based performance model for indoor mmWave wearable communication. In: 2016 IEEE International Conference on Communications (ICC), pp. 1–6. Malaysia (2016)
19. Jeffrey, G.A., et al.: Modeling and analyzing millimeter wave cellular systems. IEEE Trans. Commun. **65**(1), 403–430 (2017)
20. Theodore, S.R., et al.: Millimeter wave mobile communications for 5G cellular: it will work!. IEEE Access **1**, 335–349 (2013)

Joint Sparse-AR Model Based OFDM Compressed Sensing Time-Varying Channel Estimation

Shanshan Li[1,2(✉)], Kangyong You[1], Yueliang Liu[1], and Wenbin Guo[1,2]

[1] School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
lishanshan@bupt.edu.cn

[2] Science and Technology on Information Transmission and Dissemination in Communication Networks Laboratory, Beijing, China

Abstract. In this paper, a time-varying channel estimation method based on compressed sensing (CS) is studied to reduce the pilot overhead for orthogonal frequency division multiplexing (OFDM) system. By taking advantage of the dynamic characteristic and temporal correlation of time-varying channel, we propose a novel channel estimation scheme based on joint sparse-autoregressive (AR) model. The proposed method performs the following two steps in a sliding window strategy. Firstly, the channel delay structure is estimated using the proposed sparsity adaptive simultaneous orthogonal matching pursuit (SASOMP) algorithm. Secondly, with the channel delay estimation, a reduced order Kalman filter (KF) is performed to obtain the amplitude of channel. Simulation results indicate that the proposed method is capable of recovering the time-varying channel with much lower pilot overhead than conventional CS-based channel estimators with a superior estimation performance.

Keywords: Channel estimation · Orthogonal frequency division multiplexing · Compressed sensing · Joint sparse-AR model

1 Introduction

Orthogonal Frequency Division Multiplexing (OFDM) has been widely used in many merging communication standards such as IEEE802.16e (WIMAX), LTE/LTE-Advanced, and next generation [1–3] due to its robustness against frequency selective fading channels and high spectral efficiency. Accurate channel state information (CSI) is crucial for the coherent detection in OFDM systems. Many pilot-aided channel estimation (PACE) methods such as the least squares (LS) [4] and the linear minimum mean-squared error (LMMSE) [5] have been studied. Since these linear methods do not consider the sparse feature of channel, they usually require large pilot overhead, leading to low spectral efficiency.

It is known that the wireless channel models exhibit sparse feature as the communication bandwidth and sampling rate increase [6]. In the context of

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 762–771, 2019.
https://doi.org/10.1007/978-981-13-6264-4_90

channel estimation for OFDM systems, spectral efficient techniques that leverage this sparsity using CS theory [7–9] have been proposed. However, these above methods mainly concentrate on block fading channel model, where the number of propagation paths (sparsity) and path delays (sparsity structure) are usually assumed fixed, which can not reflect the dynamic characteristic in practical environment. In addition, it is known that the first order AR model [10] accurately captures the local behavior of fading wireless channels. In order to exploit both the sparse feature and the dynamic feature, channel estimation methods combining the orthogonal matching pursuit (OMP) algorithm with AR model have been investigated in [11,12]. However, these methods require channel sparsity as one of their inputs, while the channel sparsity cannot be obtained directly because the channel delay structure keeps changing in time-varying channels. Moreover, CS-based channel estimators proposed in [11,12] ignore the temporal correlation of sparse structure. Further studies [13,14] have shown that the path delays usually vary much slower than the path gains, which indicates the temporal correlation of delay structure among adjacent OFDM durations.

In this paper, a novel channel estimation scheme based on joint sparse-AR model is proposed in order to reduce the large pilot overhead for time-varying channel estimation in OFDM systems. The proposed method takes both the dynamic characteristic and the temporal correlation of sparsity structure into account and contains two steps. First, based on the joint sparse feature of channels, we estimate the channel delay structure by using the proposed sparsity adaptive simultaneous orthogonal matching pursuit (SASOMP) algorithm, which can obtain the channel sparsity adaptively. Secondly, utilizing the state transition characteristics of channel, a reduced order KF is performed to estimate the channel amplitude. In addition, the above two-stage process is based on a sliding window strategy, which ensures the estimation accuracy and improves the adaptability to real-time OFDM symbol streams. Simulation results show the superiority of the proposed method in terms of pilot overhead, normalized mean square error (NMSE), and bit error rate (BER).

The remainder of this paper is organized as follows. Section 2 introduces the system model and signal model over a time-varying channel. Section 3 describes the formulation of the proposed joint sparse-AR model and channel estimation scheme. In Sect. 4, simulation results are presented to demonstrate the superior performance of the proposed scheme and followed by the conclusion in Sect. 5.

2 System Model and Problem Formulation

2.1 Channel Model

The CIR of the time-varying channel can be denoted by

$$h(t,\tau) = \sum_{k=0}^{K(t)-1} a_k(t)\,\delta(\tau - \tau_k(t)) \tag{1}$$

where $K(t)$, $a_k(t)$ and $\tau_k(t)$ denote the path number, the fading coefficient and channel delay spread of the kth path, respectively. The fading coefficient can be represented as $a_k(t) = \alpha_k e^{2\pi f_{d,k} t}$, where α_k and $f_{d,k}$ represent the magnitude and Doppler shift (Hz) of kth path.

We concentrate on the slowly time-varying channel scenario, where the channel side information (i.e., multipath number $K(t)$ and delay structure $\tau_k(t)$) remains unchanged among a block of J OFDM symbols, and the channel gains change slowly over adjacent OFDM symbols. It is a reasonable assumption in most wideband wireless communication systems as the rate of channel varying is usually slower than that of data transmission. Then, in a symbol block, we can rewrite the channel model in (1) as

$$h(t,\tau) = \sum_{k=0}^{K-1} a_k(t)\,\delta(\tau - \tau_k) \tag{2}$$

Since the system sampling interval T_s is usually much smaller than channel delays τ_k, the channel coefficient of the lth path h_l has nonzero value only when $lT_s = \tau_k$. The discrete channel vector $\mathbf{h} = (h_0, h_1, \ldots, h_{L-1})^T$ is a K-sparse vector. Note that, since the channel structure keeps changing, the sparsity is unknown to the receiver directly.

2.2 Signal Model

Considering an OFDM system that consists of N subcarriers, where P subcarriers are reserved for pilot tones. Fig. 1 shows the pilot pattern used in our system. The pilots are allocated at random locations over N subcarriers and different among adjacent OFDM symbols.

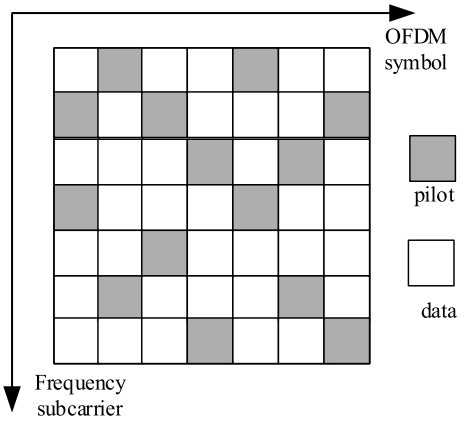


Fig. 1. The pilot pattern

Assumed that the CP length of each OFDM symbol is no less than the delay spread such that there is no inter-symbol interference (ISI) and inter-carrier interference (ICI) introduced by multipath channels. A block of J OFDM symbols are sent into a slowly time-varying channel, and then the received signal after DFT can be denoted by

$$\begin{aligned}\mathbf{Y}_n &= FFT\left(IFFT\left(\mathbf{X}_n\right) \otimes \mathbf{h}_n + \mathbf{w}_n\right) \\ &= diag\left(\mathbf{X}_n\right)\mathbf{F}\mathbf{h}_n + \mathbf{W}_n\end{aligned} \tag{3}$$

where n represents the nth OFDM symbol, $\mathbf{Y}_n \in \mathbb{C}^{N\times 1}$, $\mathbf{X}_n \in \mathbb{C}^{N\times 1}$, $\mathbf{w}_n \in \mathbb{C}^{N\times 1}$ and $\mathbf{W}_n \in \mathbb{C}^{N\times 1}$ denote the transmitted signal in frequency domain, received signal in frequency domain, complex white Gaussian noise in time and in frequency domain, respectively. The noise has zero mean and σ_w^2 variance. $\mathbf{F}_n \in \mathbb{C}^{N\times L}$ is the DFT matrix. The received signal at pilot locations can be represented as

$$\mathbf{Y}_{P,n} = diag\left(\mathbf{X}_{P,n}\right)\mathbf{F}_P\mathbf{h}_n + \mathbf{W}_{P,n} \tag{4}$$

where $\mathbf{X}_{P,n} \in \mathbb{C}^{P\times 1}$, $\mathbf{Y}_{P,n} \in \mathbb{C}^{P\times 1}$ and $\mathbf{W}_{P,n} \in \mathbb{C}^{P\times 1}$ denote the transmitted signal, received signal and noise at pilot locations, respectively. $\mathbf{F}_P \in \mathbb{C}^{P\times L}$ is the partial matrix of $\mathbf{F}$ after selecting P rows of pilot subcarriers.

Since the channel vector $\mathbf{h}_n$ is a K-sparse vector, (4) can be converted into a CS reconstruction problem [7] with the sensing matrix $\boldsymbol{\Phi}_n = diag\left(\mathbf{X}_{P,n}\mathbf{F}_P\right)$ and observation vector $\mathbf{Y}_{P,n}$.

3 Proposed Channel Estimation

3.1 Joint Sparse-AR Model

As mentioned in Sect. 2, the continuous J channel vectors share a common sparse support, while the channel gains change for every symbol. Assuming that the support vector is $\boldsymbol{\theta} = [\theta_0, \theta_1, \ldots, \theta_l, \ldots, \theta_{L-1}]^T$, $\|\boldsymbol{\theta}\|_0 = K$. Further, we introduce AR model to capture the time-varying behavior of channel amplitude. The channel vector $\mathbf{h}$ can be written as

$$\mathbf{h}_n = diag\left(\boldsymbol{\theta}\right)\mathbf{s}_n \tag{5}$$

$$\mathbf{s}_n = \mathbf{B}\mathbf{s}_{n-1} + \mathbf{v}_n \tag{6}$$

where $\boldsymbol{\theta}$ is the common support of J channel vectors and $\mathbf{s}_n$ is amplitude of the nth channel vector. The state transition matrix $\mathbf{B}$ and the process noise $\mathbf{v}_n$ can be obtained through the Yule–Walker equation [11], which can be denoted by

$$\mathbf{B} = J_0\left(2\pi f_d T_{ofdm}\right)\mathbf{I} \tag{7}$$

$$\mathbf{V} = \sigma_l^2\left(1 - J_0^2\left(2\pi f_d T_{ofdm}\right)\right)\mathbf{I} \tag{8}$$

where $J_0\left(\cdot\right)$ represents the zero-order Bessel function of the first kind, f_d denotes the Doppler frequency in Hertz and T_{ofdm} denotes the OFDM symbol period. $\mathbf{I}$ is the unit matrix and σ_l^2 is power of the lth channel path.

Combining (4), (5) and (6), the joint sparse-AR model for slowly time-varying channel estimation in OFDM systems can be written as

$$\begin{cases} \mathbf{s}_n = \mathbf{B}\mathbf{s}_{n-1} + \mathbf{v}_n \\ \mathbf{Y}_{P,n} = \boldsymbol{\Phi}_n diag\left(\boldsymbol{\theta}\right)\mathbf{s}_n + \mathbf{W}_{P,n} \end{cases} \tag{9}$$

The state space model of (9) differs from the traditional AR model in the following two aspects. First, the observation equation is an underdetermined equation. Secondly, the above J signals share a common support while the sparsity is unknown. The optimization problem to solve (9) is given as

$$\underset{\mathbf{s}_n,\theta\in\{0,1\}^L}{min} \sum_{n=1}^{J} \left\|\mathbf{Y}_{P,n} - diag(\mathbf{X}_n^P)\mathbf{F}_P diag(\theta)\mathbf{s}_n\right\|_2^2 + \lambda_1 \sum_{n=1}^{J} \left\|\mathbf{s}_n - \mathbf{B}\mathbf{s}_{n-1}\right\|_2^2 + \lambda_2 \|\theta\|_1 \tag{10}$$

where λ_1 and λ_2 are the penalty coefficients.

3.2 The Proposed Channel Estimation Scheme

Observed that (10) cannot be solved directly since the integer constraint of $\boldsymbol{\theta} \in \{0,1\}^L$. In this paper, we propose a novel joint sparse-AR model based channel estimate scheme for OFDM systems, which performs in a strategy of sliding window, as Fig. 2 shows. The sliding window process ensures the estimation accuracy and improves the adaptability to real-time OFDM symbol streams.

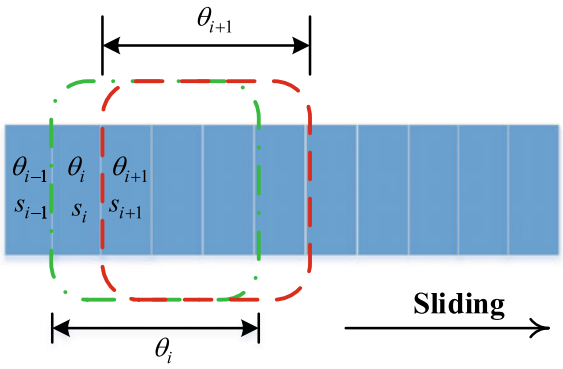


Fig. 2. The sliding window strategy

Step 1: Estimate the channel delay structure using a proposed support recovery algorithm in *Algorithm* 1, which doesn't require the channel sparsity.

This algorithm is based on the standard simultaneous orthogonal matching pursuit algorithm (SOMP) [15] but has difference in stop condition. We change the loop termination condition by identifying if the residual energy exceeds a

certain threshold V_{th}. The overall residual energy of each iteration is calculated by averaging that of J OFDM symbols. In addition, a LS-based noise energy estimate method is proposed in [16] to seek the threshold.

Step 2: Perform a reduced order KF to estimate the amplitude $\hat{\mathbf{s}}_{n|n}$ with the channel path delays estimation. In detail, the filtering process is given as:

$$\hat{\mathbf{s}}_{n|n-1} = \mathbf{B}\hat{\mathbf{s}}_{n-1|n-1} \tag{11}$$

$$\mathbf{P}_{n|n-1} = \mathbf{B}\mathbf{P}_{n-1|n-1}\mathbf{B}^H + \mathbf{V} \tag{12}$$

$$\mathbf{K}_n = \mathbf{P}_{n|n-1} diag\left(\hat{\boldsymbol{\theta}}\right) \boldsymbol{\Phi}_n^H \left(\boldsymbol{\Phi}_n diag\left(\hat{\boldsymbol{\theta}}\right) \mathbf{P}_{n|n-1} diag\left(\hat{\boldsymbol{\theta}}\right) \boldsymbol{\Phi}_n^H + \sigma_w^2 \mathbf{I}\right)^{-1} \tag{13}$$

$$\hat{\mathbf{s}}_{n|n} = \hat{\mathbf{s}}_{n|n-1} + \mathbf{K}_n \left(\mathbf{Y}_{P,n} - \boldsymbol{\Phi}_n diag\left(\hat{\boldsymbol{\theta}}\right) \hat{\mathbf{s}}_{n|n-1}\right) \tag{14}$$

$$\mathbf{P}_{n|n} = \left(\mathbf{I} - \mathbf{K}_n \boldsymbol{\Phi}_n diag\left(\hat{\boldsymbol{\theta}}\right)\right) \mathbf{P}_{n|n-1} \tag{15}$$

where $\hat{\mathbf{s}}_{n|n}$ is the amplitude estimate of n_{th} OFDM symbol duration. $\mathbf{V}$ and $\boldsymbol{\sigma}_w^2 I$ represent the process noise matrix and the observation noise matrix. $\mathbf{K}_n$ and $\mathbf{P}_{n|n}$, etc. have their usual meanings as in the KF literature. The above KF equations are initialized by setting $\hat{\mathbf{s}}_{0|0} = \mathbf{0}_{L\times 1}$, $\mathbf{P}_{0|0} = \mathbf{I}_{N\times N}$.

Finally, the nth estimated channel vector $\hat{\mathbf{h}}_n$ is denoted by

$$\hat{\mathbf{h}}_n = diag\left(\hat{\boldsymbol{\theta}}\right) \hat{\mathbf{s}}_{n|n}. \tag{16}$$

Algorithm 1. SASOMP

1: Input: Observation signal vector $\mathbf{Y}_{P,n}$, sensing matrix $\boldsymbol{\Phi}_n$, threshold V_{th}
2: Initialization the residual $\mathbf{r}_n = \mathbf{Y}_{P,n}$, the common support set $\hat{\boldsymbol{\theta}} = \varnothing$, the residual energy $V_e = \frac{1}{T}\sum_{n=1}^{J} \|\mathbf{Y}_{p,n}\|_2^2$
3: **while** $V_e > V_{th}$ **do**
4: Find an index $\hat{l}$ that solves the optimization problem
$$\hat{l} = \max_{l\in\{1,\ldots,L\}} \sum_{n=1}^{J} \left\|\boldsymbol{\Phi}_n^H(l)\, \boldsymbol{r}_n\right\|_2^2$$
We use $\boldsymbol{\Phi}_n(l)$ to denote the lth column vector of sensing matrix $\boldsymbol{\Phi}_n$
5: Update the support set $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} \cup \hat{l}$
6: Solve least-square problem and calculate new residual:
$$\hat{\mathbf{h}}_n = \left[\boldsymbol{\Phi}_n^{\boldsymbol{\theta}}\right]^{\dagger} \mathbf{Y}_{P,n}, \qquad \mathbf{r}_n = \mathbf{Y}_{P,n} - \boldsymbol{\Phi}_n^{\boldsymbol{\theta}} \hat{\mathbf{h}}_n$$
We use $\boldsymbol{\Phi}_n^{\hat{\theta}}$ to denote the partial matrix of $\boldsymbol{\Phi}_n$ by selecting rows corresponding to nonzero value of $\hat{\boldsymbol{\theta}}$
7: Calculate the new residual energy:$V_e = \frac{1}{T}\sum_{n=1}^{J} \|\mathbf{r}_n\|_2^2$
8: **end while**
9: return $\hat{\boldsymbol{\theta}}$

4 Simulation Results

In this section, we demonstrate the performance of the proposed channel estimation scheme through Monte Carlo simulations. We consider the parameters in the 3GPP/LTE broadband standard [17]. We use an OFDM system with 256 subcarriers. Each subcarrier occupies 15 kHz, resulting in an OFDM symbol duration of 83.3us with Cyclic Prefix (CP) of 16.67us (64 subcarriers). The length of discrete channel (L) is set equal to the CP length. Each OFDM frame consists of $J = 7$ OFDM symbols. The data is transmitted using a (2, 1, 7) convolutional code with QPSK modulation. We use the pilot pattern in Fig. 1 for CS-based methods and equally spaced pilots for LS channel estimator. The pilot number is 8 when not specified. The complex Rayleigh fading channel is generated using Jake's model for given Doppler frequency of 24 Hz, which corresponds to a slowly time-varying channel.

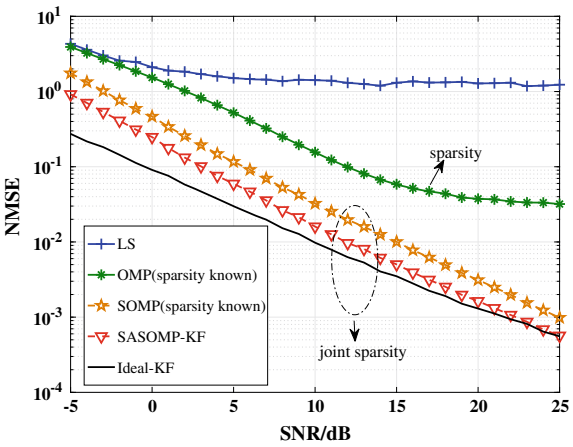


Fig. 3. NMSE performance of different schemes in a slowly time-varying channel, compared to the support ideal-KF channel estimator as a function of SNR in dB

The proposed channel estimation scheme is evaluated from three aspects: normalized mean square error (NMSE) performance, BER performance, and pilot overhead. Since the path delays estimation has a great influence on the filtering accuracy, the NMSE performance obtained with perfect knowledge of the path delays (labeled as Ideal-KF) is given in Fig. 2 as a reference. The coded performance of a receiver with perfect knowledge of the channel is shown in Fig. 3. The proposed method is labeled as SASOMP-KF in simulation results.

$$\rho = 10lg\frac{E\left[\sum \|\mathbf{h}_n - \mathbf{h}_{n,est}\|^2\right]}{E\left[\sum \|\mathbf{h}_n\|^2\right]} \tag{17}$$

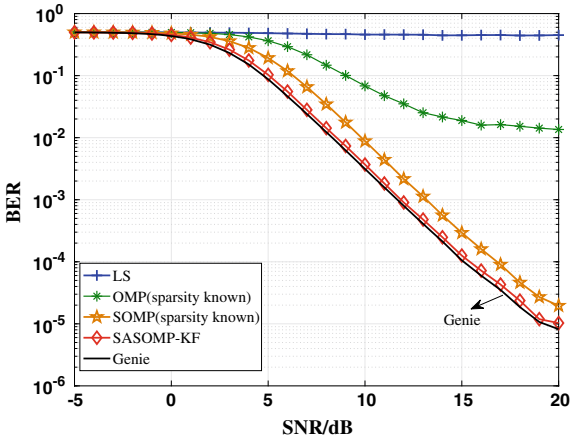


Fig. 4. BER performance of different schemes in a slowly time-varying channel as a function of SNR in dB

As shown in Fig. 3, among the four channel estimators, three kinds of CS-based channel estimators outperform the conventional LS channel estimator. Moreover, among three kinds of CS channel estimator, the proposed channel estimation scheme is better than OMP and SOMP from two aspects. On the one hand, SASOMP-KF has better NMSE performance, which is very close to Ideal-KF. Since we consider both the dynamic feature and the temporal correlation of sparsity structure of channel. On the other hand, SASOMP-KF scheme uses the proposed SASOMP algorithm to estimate the path number adaptively while

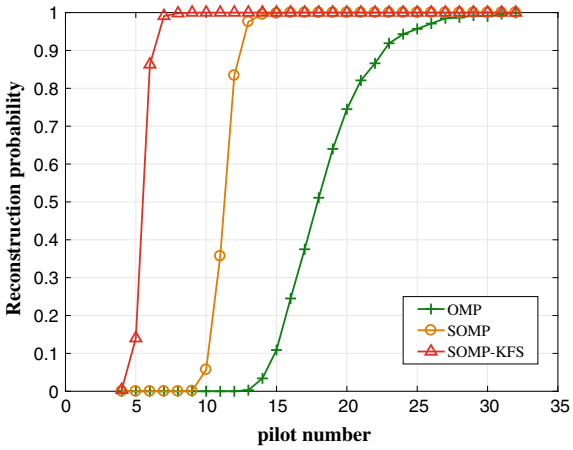


Fig. 5. The relationship of reconstruction probability and the pilots number, with $f_d = 24\,Hz$ and SNR $= 20$ dB

OMP and SOMP need sparsity as their input parameter, which indicates the effectiveness and convenience of SASOMP-KF scheme in practical application.

We depict the BER performance of the proposed algorithm in Fig. 4. While the SASOMP-KF performs about 1.5 dB better than the SOMP algorithm, it is only a fraction of dB away from performance of the genie receiver, which has perfect channel knowledge. The OMP algorithm meets its error floor when SNR>15 dB, which indicates its poor BER performance even in coded system. The LS channel estimator presents the worst BER performance under the low pilot overhead condition.

The comparison in pilot overhead of three CS-based channel estimators when SNR is 20 dB is showed in Fig. 5. The pilot overhead is defined as the percentage of pilot number over the total subcarriers when channel is recovered successfully. Assume that channel is recovered successfully when $\rho < 0.005$. From Fig. 5, the comparison of pilot overhead is as follows: SASOMP-KF (3.125%)<SOMP (6.25%)<OMP (12.5%). The overhead of SASOMP-KF is only 1/4 of OMP and 1/2 of SOMP. It indicates that the proposed SASOMP-KF scheme can further reduce the pilot overhead of conventional CS-based channel estimators.

5 Conclusion

In this paper, we consider both the dynamic characteristic and the temporal correlation of channel and propose a channel estimation scheme based on the joint sparse-AR model. The method performs in a strategy of sliding window, which ensures the estimation accuracy and improves the adaptability to real-time OFDM symbol streams. First, in order to estimate the channel delay structure, the SASOMP algorithm is proposed, which can obtain the channel sparsity adaptively. Second, a reduced order KF algorithm is used to acquire the accurate estimation of channel amplitude. Simulation results show that the proposed method outperforms the conventional CS-based channel estimators in terms of NMSE, BER, and pilot overhead. In our simulation scenario, the pilot overhead of proposed method is as low as 3.125%, which can greatly increase the system spectral efficiency.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant No. 61271181 and 61571054) and the Project Funded by Science and Technology on Information Transmission and Dissemination in Communication Networks Laboratory.

References

1. Hao, J., Wang, J., Wu, Y.: A new equalizer in doubly-selective channels for TDS-OFDM. IEEE Trans. Broadcast. **61**(1), 91–97 (2015)
2. Dahlman, E., Parkvall, S., Skld, J., et al.: 4G-LTE/LTE-advanced for mobile broadband **21**(3), iv (2011)

3. Zheng, K., Zhao, L., Mei, J., Dohler, M., Xiang, W., Peng, Y.: 10 Gb/s hetsnets with millimeter-wave communications: access and networking-challenges and protocols. IEEE Commun. Mag. **53**(1), 222–231 (2015)
4. Lee, Y.S., Bang, Y.J., Lee, J.H., Park, Y.O.: LS channel estimation performance analysis for RoF channel environment in the OFDM system. In: Proceedings of the COIN, pp. 1–2. Jeju (2014)
5. Hsieh, C.Y., Lin, D.W., Ma, C.: LMMSE-based channel estimation for LTE-advanced MIMO downlink employing UE-specific reference signals. In: Proceedings of the IEEE VTC, pp. 1–5. Glasgow (2015)
6. Kannu, A.P., Schniter, P.: On communication over unknown sparse frequency-selective block-fading channels. IEEE Trans. Inf. Theory **57**(10), 6619–6632 (2011)
7. Bajwa, W.U., Haupt, J., Sayeed, A.M., Nowak, R.: Compressed channel sensing: a new approach to estimating sparse multipath channels. Proc. IEEE **98**(6), 1058–1076 (2010)
8. Son, N.T., Guo, S., Chen, H.: Impact of channel models on compressed sensing recovery algorithms-based ultra-wideband channel estimation. IET Commun. **7**(13), 1322–1330 (2013)
9. Berger, C.R., Wang, Z., Huang, J., Zhou, S.: Application of compressive sensing to sparse channel estimation. IEEE Commun. Mag. **48**(11), 164–174 (2010)
10. Feng, T., Field, T.R., Haykin, S.: Stochastic differential equation theory applied to wireless channels. IEEE Trans. Commun. **55**(8), 1478–1483 (2007)
11. Chen, B., Cui, Q., Yang, F., Xu, J.: A novel channel estimation method based on Kalman filter compressed sensing for time-varying OFDM system. In: Proceedings of the IEEE WCSP, pp. 1–5. Hefei (2014)
12. Wang, D., Niu, K., He, Z., Tian, B.: Pilot-aided channel estimation method based on compressed sensing and Kalman filtering in OFDM systems. In: Proceedings of the IEEE ICWITS, pp. 1–4. Honolulu, HI (2010)
13. Gao, Z., Zhang, C., Wang, Z., Chen, S.: Priori-information aided iterative hard threshold: a low-complexity high-accuracy compressive sensing based channel estimation for TDS-OFDM. IEEE Trans. Wirel. Commun. **14**(1), 242–251 (2015)
14. Dai, L., Wang, J., Wang, Z., Tsiaflakis, P., Moonen, M.: Spectrum- and Energy-Efficient OFDM based on simultaneous multi-channel reconstruction. IEEE Trans. Signal Process. **61**(23), 6047–6059 (2013)
15. Tropp, J., Gilbert, A.C., Strauss, M.J.: Simulataneous sparse approximation via greedy pursuit. In: Proceedings of the ICASSP, Philadelphia (2005)
16. Xie, H., Andrieux, G., Wang, Y., Diouris, J.F., Feng, S.: Threshold based most significant taps detection for sparse channel estimation in OFDM system. In: Proceedings of the IEEE ICSPCC, pp. 1–5. KunMing (2013)
17. Zyren, J.: Overview of the 3GPP long term evolution physical layer. Freescale Semiconductor (2007)

An Improved LEACH Algorithm Based on Mobile Nodes

Xiang Liu(✉), Jia Song Mu, and Xiang Dong Yi

Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin, China
793448276@qq.com

Abstract. Many previously proposed LEACH protocols do not care about the surplus energy of the mobile node and the environmental positional relationship between the mobile nodes and the settled sink node when selecting the boss of the cluster node. The newly proposed algorithm first considers how to judge the variance of the surplus energy of the mobile node in the topological network and the average surplus energy of the remaining nodes, then a different threshold formula is selected by judging the length between the boss of the cluster node and the fixed sink node, and finally, a satisfactory cluster leader node is confirmed to transfer messages with the sink node through this series of operations. The simulation consequences of MATLAB indicate that the improved LEACH protocol algorithm prolongs the network lifetime, improves the node's data transmission capability, and reduces the node's death rate.

Keywords: LEACH protocol · Cluster head selection · Variance · Threshold

1 Introduction

Wireless sensor networks are now permeating every corner of our lives and participating in our lives [1–3]. In recent years, with the widespread recognition of the theory of routing protocols, algorithms based on LEACH have emerged in an endless stream. However, in LEACH protocol and its numerous improvements, it is rarely considered that a suitable cluster head node is confirmed by judging variance of the surplus energy of the mobile node and the medial surplus energy of network remaining nodes and the length of the range between the mobile node and the settled sink node in the dynamic simulation environment. For these two sides, an improved LEACH algorithm is proposed and verified by the simulation experiments.

2 LEACH Protocol Description

The LEACH [4] algorithm is a clustering routing algorithm that is different from plane routing. It has its own uniqueness and particularity. The explanation of the members' leader in the cluster selection stipulation in the LEACH protocol is: nodes generate

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 772–779, 2019.
https://doi.org/10.1007/978-981-13-6264-4_91

random values from 0 to 1 that is compared with the computable threshold $T(n)$ in the topological simulation network. The leader of the members in the cluster is the node that selects the small threshold and broadcasts its message.

3 Improved LEACH Algorithm Description

3.1 Comparison of Residual Energy and Variance of Nodes

It simply compares the random number with the threshold and proposes a new and improved optimization algorithm that discusses the importance of surplus energy in mobile nodes. First, calculate the variance of the surplus energy of each surviving mobile node in the current circulation network and the average surplus energy of the remaining mobile node. Then, compare the variance with the set specific value. If the variance is greater than the preset specific value, perform 2.2. The detailed equation is as follows:

$$E_a = E_t \times \left(1 - \frac{r}{r_m}\right)/n \tag{1}$$

$$\mathrm{B} = \left(\sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{E_c} - \mathrm{E_a})^2\right)/\mathrm{S} \tag{2}$$

$$B > F \tag{3}$$

Among them, E_c is the surplus energy of each surviving mobile node in the current circulation; E_a is the medial surplus energy of the remaining nodes in the current round network; E_t is global energy of the network before the simulation topology begins; S is the value of surviving nodes minus dead nodes in the current topological network; $F = 0.008$; $n = 100$; r_m is the maximum value of simulations for this algorithm; B is the variance.

3.2 Choice of Node Position and Threshold Formula

After the variance is not less than the specified value in 2.1, the data transmission length between the mobile nodes and the sink node in topology simulation is first calculated, and the appropriate threshold formula is selected by comparing the different ranges of the distance with the random number from 0 to 1, and finally, the optimal the boss of the cluster node is selected to extend the network's operating cycle. The improved threshold formula based on weighted processing is presented below:

$$\mathrm{t(n)} = \begin{cases} \frac{\mathrm{P_i}}{1-\mathrm{P_i}*(\mathrm{r}*\mathrm{mod}(\mathrm{P_i}))} \times (\mathrm{A_1} * (\mathrm{E_c} * \mathrm{S}/\mathrm{E_t}) + \mathrm{A_2} * \mathrm{d}), & \text{if } \mathrm{n} \in \mathrm{Z} \\ 0, & \text{other} \end{cases} \tag{4}$$

Among them,

$$P_i = \frac{P \times n \times E_c^2}{E_t \times E_a} \tag{5}$$

Among them, P_i is the proportion of mobile nodes that survived in the current circulation are proposed as the boss of members in the cluster; P is the ratio of all boss of cluster node and all of nodes in the simulation topology; the length between mobile nodes in the dynamic network is divided into three ranges of 20, 60, and 100, where A_1 and A_2 are based on different distances. d is the length between each surviving mobile node and the sink node in the topology diagram; Z is the number of mobile nodes that are not proposed as the boss of members in the cluster among the mobile nodes that survived the current circulation.

3.3 Energy Calculation of the Improved Algorithm

The three hypothetical conditions in the improved algorithm are as follows:

(1) Nodes and sinks can communicate with each other, and the nodes are not fixed.
(2) All nodes have finite, equal initial energy, and their own energy can be monitored anytime.
(3) Symmetric wireless channel, the energy consumed to send and receive data between mobile nodes is the same.

By adopting the assumptions of 1.2.3, the formula of energy expended by a mobile node in the network to send k-bit data is

$$E_{send} = \begin{cases} k \times \left(E_{tx} + E_{tfs} \times \text{distance}^2\right), & d \le d_0 \\ k \times \left(E_{tx} + E_{tmp} \times \text{distance}^4\right), & \text{otherwise} \end{cases} \tag{6}$$

In this formula, distance is the length of data transmission between the mobile node and other mobile nodes, E_{tx} is the energy expended by the sensor to transmit 1-bit data, and E_{tfs} is the energy generated by sensor nodes to deliver 1-bit data per unit area under the disengaged roomage channel model.E_{tmp} is the energy consumption coefficient under Two-ray ground reflection model [5].

From the Formula (6), the critical value of d_0 can be obtained directly:

When the formulas in (6) are equal to the following equation, we can calculate $d_0 = \sqrt{\frac{E_{tfs}}{E_{tmp}}}$.

The model of resources expended by the cluster leader node for receiving and merging m bit data is

$$E_{rm} = m \times (E_{rx} + E_{dx}) \tag{7}$$

In the equation, E_{rx} is the energy expended by the node receiving each processing of 1-bit data. E_{dx} is the energy consumed by data fusion of 1-bit data.

3.4 The General Flowchart Description

The detailed and complete description of the algorithm is as follows (Fig. 1).

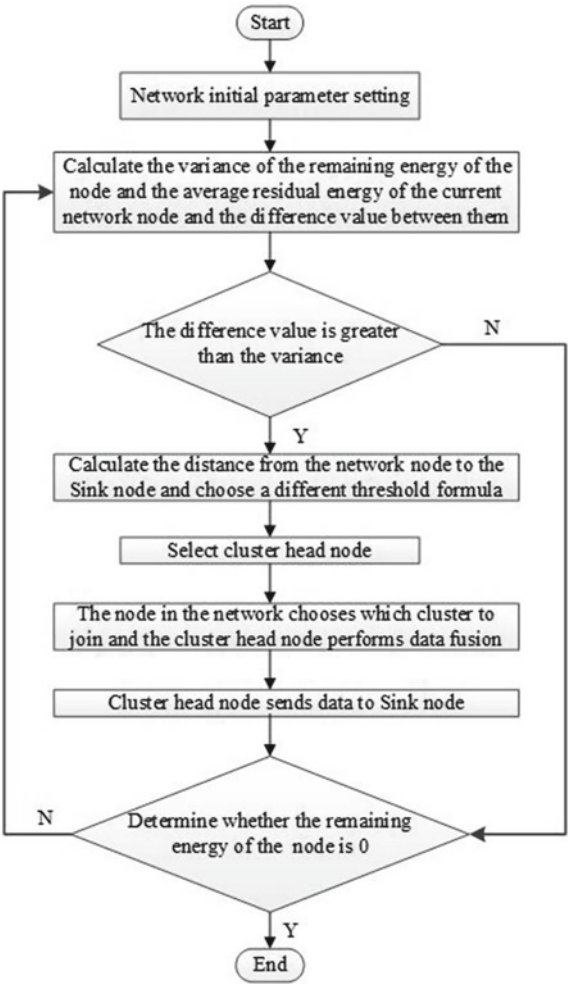


Fig. 1. Flowchart of improved LEACH algorithm

4 Simulations Performance Under the Algorithm

4.1 Model and Parameter Initialization

The amount of identical surviving mobile nodes in the simulation topology is set to 100, arbitrarily created in the 200 m × 200 m definitive area, and the sink node is fixed at the (100,100) point. The simulation parameter settings are: the cluster head optimization ratio is 0.05, the incipient energy of the mobile node just put into the topological network is 0.1 * (1 + rand) J, the resource consumption sent and received by each node is 50 nJ/bit, the energy consumption generated by nodes' data fusion is 50 nJ/bit, the system sends the data of energy consumption of 10 $nJ/bit/m^2$, the power consumption of the power amplifier in the network is 0.0013 $pJ/bit/m^4$, the amount of analog transmission information is 4000 bits, and the maximum number of simulations in the network is 1000 times.

4.2 Simulation Results and Analysis

Figures 2, 3, and 4 show performance comparisons for the improved LEACH algorithm and LEACH. The improved LEACH algorithm not only explores the variance of

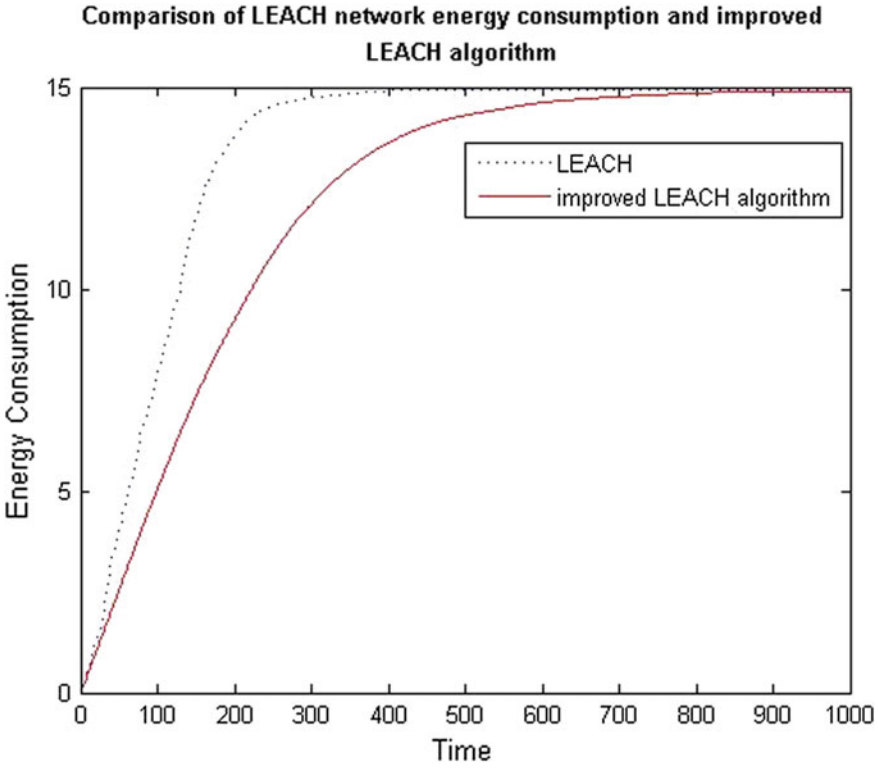


Fig. 2. Comparison of LEACH network energy consumption and improved LEACH algorithm

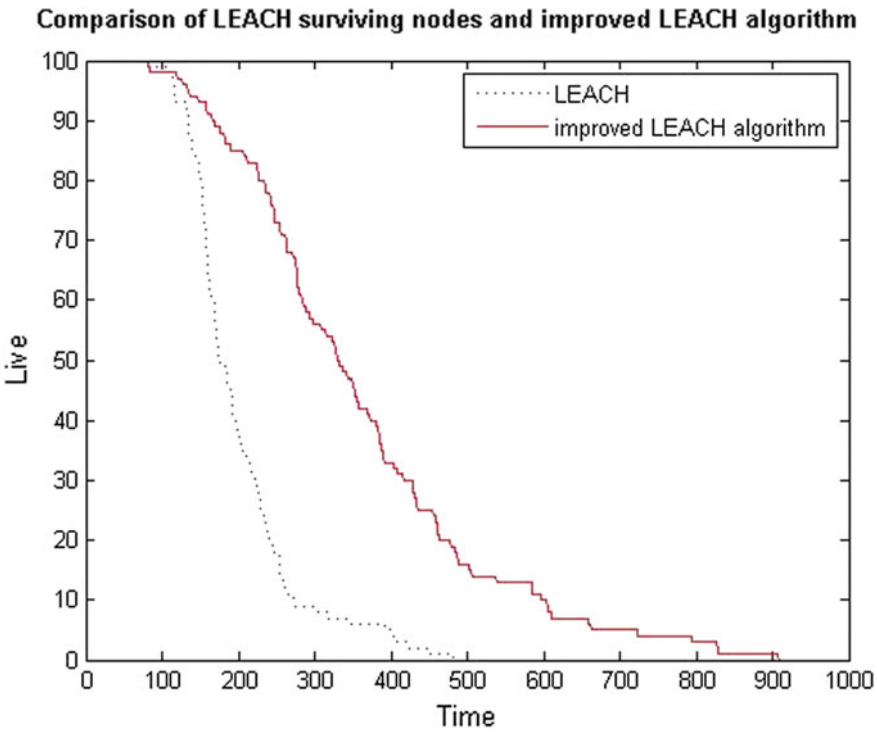


Fig. 3. Mobile node survival performance under two kinds of algorithms

the surplus energy of the surviving mobile nodes and its average surplus energy in the current topological emulation, but also inspects the positional relationship between the mobile nodes and the fixed sink nodes, so as to poise the power dissipation of the mobile nodes. The lifespan of the topological network operation and the ability to transmit data are improved in the network.

Figure 2 shows the comparison of network energy costs changes of the two algorithms. The improved LEACH algorithm has a significantly better curve change trend than the LEACH algorithm. Around 200 rounds, the longitudinal value between the network energy consumption of the improved LEACH algorithm and the LEACH algorithm reaches a maximum, and the graph is always much higher than the LEACH algorithm before 900 rounds. Therefore, the improved LEACH algorithm greatly increases the utilization of energy resources in the dynamic topology network. Figure 3 reveals that the improved algorithm precedes the LEACH algorithm in terms of surviving mobile nodes in the dynamic topology simulation. The change of the surviving quantity of the LEACH algorithm node tends to change steadily. Only about 900 rounds ago, all the nodes died. However, the amount of mobile nodes in the LEACH algorithm is essentially zero in the 400 rounds. Figure 4 shows the comparison of the data transmission capabilities of the two algorithms. It is obvious that the LEACH algorithm achieves the maximum value of data transmission in about 400 rounds.

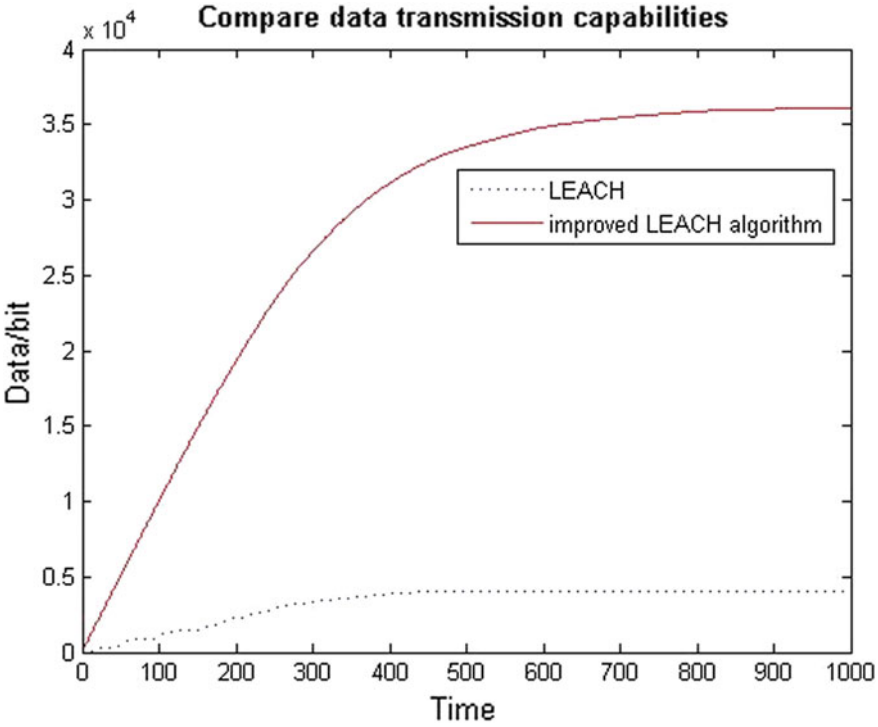


Fig. 4. Data transmission capabilities under two kinds of algorithms

The improved LEACH algorithm is approximately 7.5 times that of the LEACH algorithm in 300 rounds, and reaches the peak of data transmission around 3.6×10^4 bit in about 850 rounds. Therefore, the improved LEACH algorithm further improves the ability of node data transmission and improves the stability of the network.

5 Conclusions

In this article, regardless of network energy consumption, node survivability number and node data transmission capability, the new improved algorithm based on LEACH protocol has more outstanding performance than the traditional LEACH algorithm. The overall frame of the improved algorithm in this article is to compare the variance of the surplus power of surviving mobile nodes and its average surplus power with the preset special value in the dynamic topological network, and then use the variance and the specific value to select the node with more remaining energy, and the node with more surplus energy source is designated by comparing the above the variable value and the specific value so as to cause low residual energy node in the data transmission early death, and then made a change in formulating and selecting the threshold formula t(n). First, an appropriate threshold formula was selected by judging the scope between the mobile sensor node in the dynamic topological network and the fixed sink node

position, and the appropriate sensor node was selected ultimately. Meanwhile, the newly proposed t(n) threshold formula is weighted based on the two different weights of A_1 and A_2 on the original basis. The first item is the ratio of the sum of remaining energy of the current topological node to the general energy of dynamic circumstance. The second item is the length between surviving mobile nodes and mobile nodes in the dynamic network. Different weights are given for these two items, and the threshold formula is further optimized to improve the life span of the whole dynamic condition. The algorithm improves the probability of survival of nodes and the ability of nodes to transmit data. However, in Fig. 3, there are still phenomena of premature death of nodes and a large number of deaths after 900 rounds, so further improvements are needed in the optimization algorithm.

Acknowledgments. This paper is supported by Natural Science Foundation of China (61271411), Natural Youth Science Foundation of China (61501326,61401310). It also supported by Tianjin Research Program of Application Foundation and Advanced Technology (15JCZDJC31500), and Tianjin Science Foundation(16JCYBJC16500).

References

1. Limin, S., Jianzhong, L., Yu, C., et al.: Wireless Sensor Networks, pp. 1–25. Beijing Tsinghua Press (2005)
2. Ho, L., Moh, M., Walker, Z., et al.: A prototype on RFID and sensor networks for elder healthcare: progress report. In: Proceedings of the 2005 ACM SIGCOMM Workshop on Experimental Approaches to Wireless Network Design and Analysis, pp. 70–75. ACM, Philadelphia (2005)
3. Yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. Comput. Netw. **52**(12), 2292–2330 (2008)
4. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless micro sensor networks. In: Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, pp. 3005–3014. IEEE, Maui (2000)
5. Noury, N., Herve, T., Rialle, V., et al.: Monitoring behavior in home using a smart fall sensor and position sensors. In: Proceedings of the 1st Annual International Conference on Microtechnologies in Medicine and Biology, pp. 607–610. IEEE, Lyon (2000)

A Strategy of NOMA User Pairing by Time-Sharing in Nonoverlapping Frequency Band

Yan-Jun Liang[1], Hui Li[2(✉)], and Yu-Han Dai[1]

[1] School of Information Science and Technology, Hainan University, 570228 Haikou, China
[2] School of Aeronautics and Astronautics, Zhejiang University, 310027 Hangzhou, China
hitlihuilll2@163.com

Abstract. The Non-orthogonal Multiple Access (NOMA) technique is seen as an important technology in the future 5G networks' architecture. This paper mainly studied a new method of user pairing in NOMA communication to satisfy the signal transmission and reception. It proposes pairing of users in nonoverlapping frequency bands so that two or more edge users can be paired with one central user at the same time. We made a comparison between our new user pairing model and the traditional NOMA system model. Simulation results show that this improvement not only reduces the number of users but also increases the bandwidth utilization, even more it increases the system's ergodic sum capacity.

Keywords: 5G communications · NOMA · Nonoverlapping bands · User pairing · Ergodic capacity

1 Introduction

In recent years, the number of devices connected to existing networks has shown an explosive growth trend, and the prosperity of the Internet of Things will make this figure grow even more dramatically. With these huge numbers of devices accessing the communication networks, there are numerous applications that use network resources, which require network capacity to reach a very high level. The design and implementation of the Fifth Generation (5G) network contribute to meet the high demand of network capacity and plays an important backbone in the Internet of Things. Many different research technologies make up the 5G network framework. This has led to the emergence of many different standards and specifications for the development of 5G networks. Therefore, various multipath technologies have been proposed and researched to analyze their performance. Among various 5G wireless access technologies, the spectral efficiency of Non-orthogonal Multiple Access (NOMA) is considered to be the future choice [1].

Different users communicate simultaneously at a frequency such that their respective information signal may be superimposed on the base station which allocates different transmission power on each other in NOMA. Users with better channel gains are assigned lower power, while user signals with poorer channel gains receive higher

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 780–789, 2019.
https://doi.org/10.1007/978-981-13-6264-4_92

power [2]. Since at the same time different users use the same frequency and code source in the same group, it also causes interference to become one of the serious problems in NOMA.

The user-assigned low power with good channel gain is easily interfered by the user's high power signal with poor channel gain, so it is necessary to use Successive Interference Cancellation (SIC) to repair their signals. At the same time, users with poor gain face less interference due to their high power signals and do not need to be performed with SIC or other interference cancellation. The literature [3] makes use of this method to enable high-gain and low-gain users to use the same frequency for simultaneous communication, achieving a significant increase in the overall capacity of the system. In this case, the user is a generalization of the concept, it represents all electronic communication devices in accordance with the principles of NOMA, it may be a cellular mobile phone network, a wireless sensor network sensor, or the IOT any intelligent device.

User pairing and power allocations are the key factors to achieve performance gain in NOMA [4]. To obtain significant capacity gains, users with large differences in channel gain should adapt to each other. This means that cell center users (better channel gain) should be paired with cell edge users (poor channel gain). This user pairing strategy is so-called traditional NOMA user pairing. And, the traditional NOMA user pairing mode may cause the pairing problem of the average channel gain users located between the cell center and cell edge users [5].

Generally, it is assumed that the number of users in the cell center and the cell edge is approximately the same, that is, to say the user is evenly distributed, so that each cell center user can find and match the cell edge user in NOMA [6]. However, users are usually randomly deployed in the cellular area, so that the number of users in the cell center and the cell edge is not always the same, and this, in turn, causes the user density in the cell center and the cell edge as different. And, this results in the inability to ensure that all users from the dense area are paired. Communication services can only be provided for these users through the traditional multipath approach [7], as shown in Fig. 1. How to handle the matching problem, in this case, is the challenge that the current NOMA Institute faces.

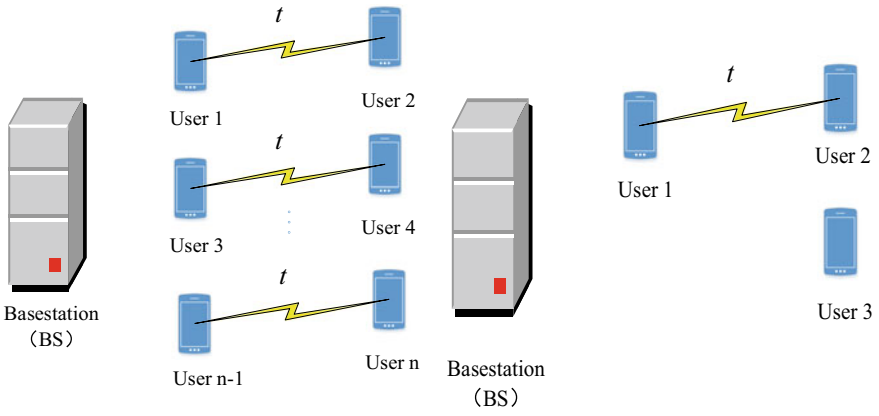


Fig. 1. User pairing strategy in traditional NOMA mode

In practice, the number of cell center users is generally smaller than that of cell edge users. Therefore, it is considered to pair one cell center user with two or even more than two cell edge users on nonoverlapping frequency bands in order to achieve a higher capacity gain than the traditional NOMA system [8]. It needs to be taken into account that in this particular case, when a cell center user is paired with multiple users at the same time, multiple SICs need to be performed, and that increases the computational complexity of information processing. However, this problem does not occur when the number of cell center users is more than the number of edge users [9].

This paper investigates the user pairing scheme of nonoverlapping frequency bands that is adaptive to similar gain users in NOMA in a time-sharing manner. In this case, the number of edge users is more than that of central users, and this causes the pairing problem of NOMA systems. This method proposes the possibility of pairing in nonoverlapping frequency bands. It will be compared with the traditional NOMA pairing simulation. Two or more similar gain users are paired with a single remote user on a time-sharing basis. Mathematical analysis was performed on the traversal and capacity of the system model. The simulation results verify the mathematical analysis and compare it with the traditional NOMA solution.

2 System Model

First, we set up a system model consisting of a cell center and two cell edge users. User 1 is located in the center of the cell, User 2 and User 3 are located at the edge of the cell, User 1 has a channel gain of $|a_1|^2$, and two edge Users 2 and 3 have channel gains of $|a_2|^2$ and $|a_3|^2$, respectively, and $|a_1|^2 > |a_2|^2 \simeq |a_3|^2$. The channel coefficients on each link are $a_1 \sim CN(0, \lambda_1 = l_1^{-v})$, $a_2 \sim CN(0, \lambda_2 = l_2^{-v})$ 和$a_3 \sim CN(0, \lambda_3 = l_3^{-v})$, l_1, l_2, and l_3 represent the link distance from the base station to the User 1, the User 2, and the User 3, respectively, as shown in Fig. 1. We use v to represent the path loss index.

According to the NOMA pairing principle, if the channel gains of two or more users differ greatly, they can be paired, grouped, and multiplexed at the same time-frequency resource [10]. As can be seen from Fig. 1, the channel gains between User 1 and User 2, User 3 are very different, which means that User 1 can be paired with User 2 or User 3. However, the channel gains of Users 2 and 3 are almost similar because they are all the edge of the cell users and very close to each other, and also they are the same distance from the base station. Therefore, the two users cannot be accommodated together in the user's center because their signals will cause great interference to each other. In other words, User 1, User 2, and User 3 cannot perform joint pairing at a common time and frequency.

Since the channel conditions of User 2 and User 3 are almost the same, as is known, in the conventional NOMA pairing, User 1 can be paired with User 2 or User 3, but cannot be paired at the same time. The unpaired user can only use the Orthogonal Frequency Division Multiplexing (OFDM) service. This limits the overall capacity of the NOMA system. In order to utilize the frequency band of unpaired users to increase the overall capacity of the system, User 1 is proposed to be paired with User 2 and User 3 at the same time in nonoverlapping frequency bands. In this way, User 1 can utilize the frequency bands of both User 2 and User 3, further increasing the overall capacity

of the system. The interference between User 2 and User 3 is further reduced. It can also pair users by means of time-sharing to increase NOMA's capacity revenue.

In addition, in order to reduce the computational complexity of the SIC at the receiving end, we set the total system bandwidth to be 1 Hz (B = 1 Hz). The total time is divided into two time slots which are t_1 and t_2. In the first time slot t_1, User 1 is paired with User 2 over the full bandwidth of B Hz, where $t_1 = \alpha$, cause $0 \leq \alpha \leq 1$. For the remaining time t_2 = 1-α, User 1 and User 3 will be paired over the entire B Hz bandwidth. In this way, User 1 can use the entire B Hz bandwidth for the duration of the entire communication session, while User 2 and User 3 can use full bandwidth for part of the time at times t_1 and t_2, respectively. During any particular time slot, User 1 performs SIC to decode and cancel the signal of paired cell edge User 2/User 3 to recover its own signal. In contrast, the paired cell edge User 2/User 3 does not perform any SIC and recovers its own signal by treating the User 1 signal as noise/interference. It is important to mention that User 1 performs only one SIC procedure in each slot because it only pairs with one cell edge user in that particular slot (Fig. 2).

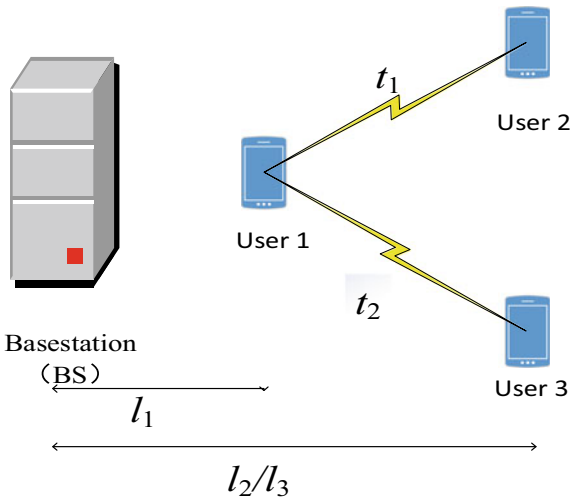


Fig. 2. A NOMA user pairing strategy that uses nonoverlapping frequency bands to enable simultaneous pairing between users

3 Ergodic Sum Capacity Analysis

In this section, we will calculate the ergodic capacity of the system. For easy calculation and analysis, the system bandwidth is set equal to be 1 Hz. The average channel gains of the two cell edge users are considered to be similar as they are almost the same distance from the BS. That is to say $|a_2|^2 \simeq |a_3|^2$. The total transmit power of the base station remains is set to be 1, P = 1, and then calculate the traversal capacity of each user.

During the time slot t_1, User 1 and User 2 are paired over the entire bandwidth B Hz. h_{1,t_1} represents the power allocated to User 1 during time slot t_1. According to NOMA's rules, the total power is $P = 1$, the power allocated to User 2 during this time slot is $h_{2,t_1} = 1 - h_{1,t_1}$. The achievable capacity of User 1 in the first time slot is as follows:

$$G_{1,t_1} = \alpha \log_2(1 + \rho |a_1|^2 h_{1,t_1}). \tag{1}$$

In the formula, ρ denotes the Signal-to-Noise Ratio (SNR) at the transmission and $\rho |a_1|^2 h_{1,t_1}$ denotes the SNR at the receiver. $X \triangleq \rho |a_1|^2 h_{1,t_1}$, the cumulative distribution function of X be given as

$$F_X(x) = 1 - e^{-\frac{x}{\lambda_1 \rho h_{1,t_1}}}. \tag{2}$$

To solve $\int_0^\infty \log_2(1+x) f_X(x) d_x$, we could get the ergodic capacity of User 1. And, as is known, $\int_0^\infty \log_2(1+x) f_X(x) d_x = \frac{1}{\ln 2} \int_0^\infty \frac{1 - F_X(x)}{1+x} d_x$, let E_i to be the exponential integral function, in that way, we get the ergodic capacity of User 1 in the first time slot as follows:

$$G_{1,t_1}^{erg} = \frac{\alpha}{\ln 2} \int_0^\infty (\frac{1}{1+x}) e^{-\frac{x}{\lambda_1 \rho h_{1,t_1}}} d_x = \frac{\alpha}{\ln 2} \{-E_i(\frac{-1}{\lambda_1 \rho h_{1,t_1}})\}. \tag{3}$$

User 1 and User 3 are paired with each other in the second time slot t_2, h_{1,t_2} and h_{3,t_2} represent the power allocated to User 1 and User 3 at this time slot, respectively. Then, the ergodic capacity of User 1 at time slot t_2 can be calculated as

$$G_{1,t_2}^{erg} = \frac{1 - \alpha}{\ln 2} \{-E_i(\frac{-1}{\lambda_1 \rho h_{1,t_2}}) e^{\frac{1}{\rho \lambda_1 h_{1,t_2}}}\}. \tag{4}$$

Then, the total ergodic capacity of User 1 can easily be got as follows:

$$G_{1,t}^{erg} = G_{1,t_1}^{erg} + G_{1,t_2}^{erg} = \frac{\alpha}{\ln 2} \{-E_i(\frac{-1}{\lambda_1 \rho h_{1,t_1}}) e^{\frac{1}{\lambda_1 \rho h_{1,t_1}}}\} + \frac{1 - \alpha}{\ln 2} \{-E_i(\frac{-1}{\lambda_1 \rho h_{1,t_2}}) e^{\frac{1}{\lambda_1 \rho h_{1,t_2}}}\}. \tag{5}$$

User 2 is paired with User 1 over the total bandwidth of B Hz during the first time slot t_1, and as we know, the power allocated to User 1 at t_1 is higher to the User 2, so the SIC does not need to be performed. Even more User 2 could decode his own signal by treating User 1's low power signal as noise. The SNR at the receiver can be used to calculate the rate that User 2 will achieve. The User 2's achievable rate is given by

$$G_{1,t_2} = \alpha \log_2(1 + \frac{\rho |a_2|^2 h_{2,t_1}}{\rho |a_2|^2 h_{2,t_1} + 1}). \tag{6}$$

Let $Y \triangleq \frac{\rho |a_2|^2 h_{2,t_1}}{\rho |a_2|^2 h_{2,t_1}+1}$, and the cumulative distribution function of Y is as follows:

$$F_Y(y) = 1 - (1 + \frac{\rho\lambda_2 h_{1,t_1} y}{\rho\lambda_2 h_{2,t_1}})^{-1} e^{-\frac{y}{\rho\lambda_2 h_{2,t_1}}}. \tag{7}$$

As $\int_0^\infty \log_2(1+y) f_Y(y) d_y = \frac{1}{\ln 2}\int_0^\infty \frac{1-F_Y(y)}{1+y} d_y$, the ergodic capacity of User 2 at time slot t_2 can be calculated as

$$\begin{aligned} G_{2,t_1}^{erg} &= \frac{\alpha}{\ln 2}\int_0^\infty \frac{1}{1+y}(\frac{\rho\lambda_2 h_{2,t_1}}{\rho\lambda_2 h_{2,t_1} + \rho\lambda_2 h_{1,t_1} y}) e^{-\frac{y}{\rho\lambda_2 h_{2,t_1}}} d_y \\ &= \frac{\alpha\lambda_2\rho h_{2,t_1}}{\ln 2(\lambda_2\rho h_{2,t_1} - \lambda_2\rho h_{1,t_1})}\{-E_i(\frac{-1}{\lambda_2\rho h_{2,t_1}}) e^{\frac{1}{\lambda_2\rho h_{2,t_1}}} + E_i(\frac{-1}{\lambda_2\rho h_{1,t_1}}) e^{\frac{1}{\lambda_2\rho h_{1,t_1}}}\}. \end{aligned} \tag{8}$$

During the second time slot t_2, User 3 is paired with User 1 over the total bandwidth. Similar to the first time slot t_1, in the second time slot t_2, $h_{1,t_2} < h_{3,t_2}$. Therefore, the User 3 does not perform any SIC process and decodes its own high power signal by treating the User 1's low power signal as noise. The data rate that User 3 can achieve is as follows:

$$G_{3,t_2} = (1-\alpha)\log_2(1 + \frac{\rho|a_3|^2 h_{3,t_2}}{\rho|a_3|^2 h_{3,t_2}+1}). \tag{9}$$

And, $Z = \frac{\rho|a_3|^2 h_{3,t_2}}{\rho|a_3|^2 h_{3,t_2}+1}$, so the ergodic capacity of User 3 is as follows:

$$G_{3,t_2}^{erg} = \frac{(1-\alpha)(\lambda_3\rho h_{3,t_2})}{\ln 2(\lambda_3\rho h_{3,t_2} - \lambda_3\rho h_{1,t_2})}\{-E_i(\frac{-1}{\lambda_3\rho h_{3,t_2}}) e^{\frac{1}{\lambda_3\rho h_{3,t_2}}} + E_i(\frac{-1}{\lambda_3\rho h_{1,t_2}}) e^{\frac{1}{\lambda_3\rho h_{1,t_2}}}\}. \tag{10}$$

In summary, it is easy to see the ergodic capacity of the whole system which could be given in

$$\begin{aligned} G_{noma}^{erg} =& G_{1,t_1}^{erg} + G_{1,t_2}^{erg} + G_{2,t_1}^{erg} + G_{3,t_2}^{erg} = \frac{\alpha}{\ln 2}\{-E_i(\frac{-1}{\lambda_1\rho h_{1,t_1}}) e^{\frac{1}{\lambda_1\rho h_{1,t_1}}}\} + \frac{1-\alpha}{\ln 2}\{-E_i(\frac{-1}{\lambda_1\rho h_{1,t_2}}) e^{\frac{1}{\lambda_1\rho h_{1,t2}}}\} \\ &+ \frac{\alpha\rho\lambda_2 h_{2,t_1}}{\ln 2(\alpha\rho\lambda_2 h_{2,t_1} - \alpha\rho\lambda_2 h_{1,t_1})}\{-E_i(\frac{-1}{\lambda_2\rho h_{2,t_1}}) e^{\frac{1}{\lambda_2\rho h_{2,t_1}}} + E_i(\frac{-1}{\lambda_2\rho h_{1,t_1}}) e^{\frac{1}{\lambda_2\rho h_{1,t_1}}}\} \\ &+ \frac{(1-\alpha)\rho\lambda_3 h_{3,t_2}}{\ln 2(\alpha\rho\lambda_3 h_{3,t_2} - \alpha\rho\lambda_3 h_{1,t_2})}\{-E_i(\frac{-1}{\lambda_3\rho h_{3,t_2}}) e^{\frac{1}{\lambda_3\rho h_{3,t_2}}} + E_i(\frac{-1}{\lambda_3\rho h_{1,t_2}}) e^{\frac{1}{\lambda_3\rho h_{1,t_2}}}\}. \end{aligned} \tag{11}$$

4 Simulation Analysis

In this section, we mainly verify the improved user-paired NOMA model by simulation and do some comparison with the traditional model. The main parameters of simulation are shown in Table 1.

Table 1. Simulation conditions

Parameter type	Parameter size
Simulation environment	Urban macro
Cell radius	500 m
CP	Normal
Scheduling mode	Proportional Fairness Scheduling (PF)
Power distribution mode	Fixed Power Allocation (FPA)
Modulation system	Quadrature Phase Shift Keying (QPSK)
Channel	Additive White Gaussian Noise (AWGN)

According to the above, the total system bandwidth and transmission power of the base station in the simulation are kept at 1, and the distance between the base station and the user is $l_1 = 0.1$, $l_2 = l_3 = 1$. User power $h_{1,t_1} = h_{1,t_2} = 0.05$, $h_{2,t_1} = h_{3,t_2} = 0.95$, $t_1 = t_2 = 0.5$. The simulation results could be seen in Fig. 3.

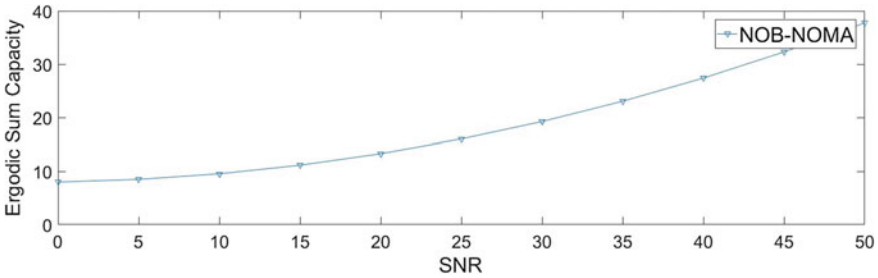


Fig. 3. Ergodic sum capacity of the NonOverlapping Frequency Band NOMA (NOB-NOMA) with different SNRs ($l_1 = 0.1$, $l_2 = l_3 = 1$, $h_{1,\,t1} = h_{1,t2} = 0.05$, $h_{2,\,t1} = h_{3,t2} = 0.95$, $t_1 = t_2 = 0.5$)

It can be understood that the ergodic capacity of the system in Fig. 3 gradually increases as the SNR increases. As the SNR becomes higher, the capacity increase ratio also increases relatively. Then, the NOMA and OFDM models in the traditional form are simulated and compared in Fig. 4.

It is easy to see that the NOMA system has higher traversal capacity at the same SNR, and there is a natural advantage over the OFDM model. Comparing with the improved NOMA system model, the following Fig. 5 is obtained.

In turn, the paired transmission signal models in the nonoverlapping band, the traditional NOMA system model, and the OFDM system model are compared to the

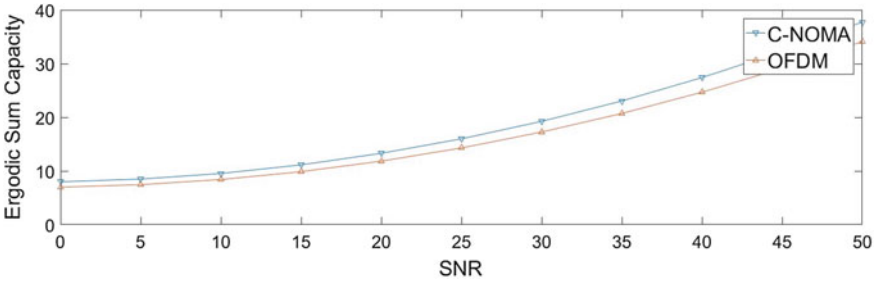


Fig. 4. Ergodic sum capacity of Conventional NOMA (C-NOMA) and OFDM with different SNRs ($l_1 = 0.1$, $l_2 = l_3 = 1$, $h_{1,t1} = h_{1,t2} = 0.05$, $h_{2,t1} = h_{3,t2} = 0.95$, $t_1 = t_2 = 0.5$)

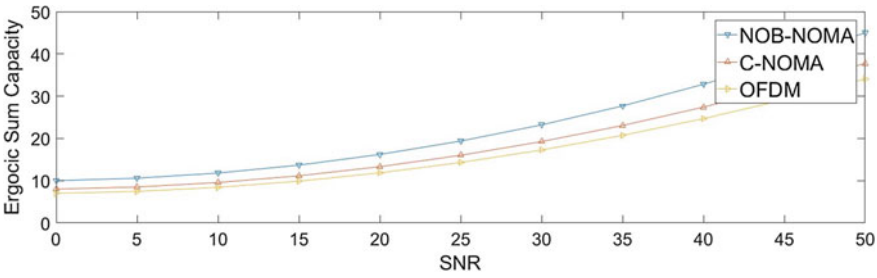


Fig. 5. Three kinds of model simulation comparison ($l_1 = 0.1$, $l_2 = l_3 = 1$, $h_{1,t1} = h_{1,\ t2} = 0.05$, $h_{2,t1} = h_{3,\ t2} = 0.95$, $t_1 = t_2 = 0.5$)

respective ergodic capacity obtained by simulation under each model. It can be clearly seen that the improved NOMA model has been improved in traversal capacity very much.

In addition, different ergodic capacity is obtained under the model considering that the relative distance between the edge user and the center user in real life is different. So, we change the distance, and compare the simulation data, the simulation results can be given in Fig. 6.

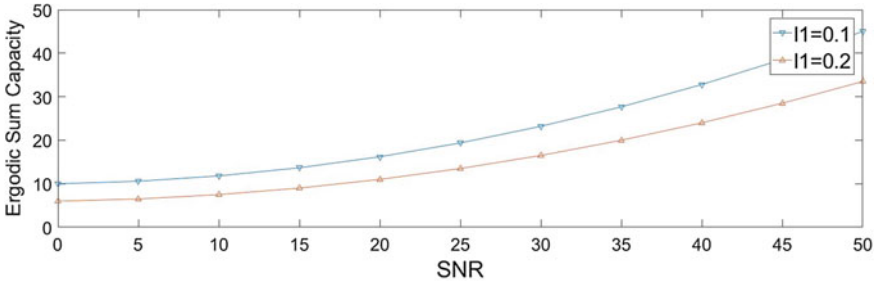


Fig. 6. Simulation comparisons with the distance changing under NOB-NOMA model

Other conditions remain unchanged, only the distance get changed, the above is when $l_1 = 0.1$, $l_2 = l_3 = 1$, below $l_1 = 0.2$, $l_2 = l_3 = 1$, you can see the system that is relatively distant from a signal with a noise-to-noise ratio of 0 which is more than twice the ergodic capacity of a system that is closer to a relative distance. With the increase of SNR, the ergodic capacity growth trend of $l_1 = 0.1$ is also faster. This proves that as the relative distance between the edge user and the central user is farther, the ergodic capacity of the paired system will increase and the growth will increase.

5 Conclusion

The traditional NOMA system pairing method often ignores the situation when the number of users in the edge and the center is inconsistent. This paper introduces a new user pairing method in NOMA, proposing that two or more users could be paired at the same time but in nonoverlapping frequency bands. It is also possible to perform a pairing transmission signal with a central user at the same time to improve the existing user pairing transmission and reception methods. The simulation results show that the improved NOMA has greatly improved ergodic capacity at the same SNR compared to the conventional NOMA system, and the ergodic capacity increases a lot with the increasing of SNR.

Some problems have also been found in the course of research. There are cases where the number of central users is more than that of marginal users, and the pairing problems of users located at relatively average positions are also the direction that we need to study in the future.

Acknowledgments. This work is supported by the Hainan Provincial Key Research and Development Program of China (ZDYF2016010 and ZDYF2018012), the National Natural Science Foundation of China (61661018), the Zhejiang Provincial Natural Science Foundation of China (LY15F010003), and the International Science and Technology Cooperation Special Project of China (2015DFR10510). Hui Li is the corresponding author of this article.

References

1. Razzaque, M.A., Milojevic-Jevric, M., Palade, A., Clarke, S.: Middleware for internet of things: a survey. IEEE Internet Things J. **3**(1), 70–95 (2016)
2. Tang, C.: Research on Signal Detection for Downlink Non-orthogonal Multiple Access of Technology. Chongqing University of Posts and Telecommunications (2016)
3. Li, A., Lan, Y., Chen, X., Jiang, H.: Non-orthogonal multiple access (NOMA) for future downlink radio access of 5G. China Commun. **12**(S), 28–37 (2015)
4. Saito, K., Benjebbour, A., Kishiyama, Y., Okumura, Y., Nakamura, T.: Performance and design of SIC receiver for downlink NOMA with open-loop SU-MIMO. In: IEEE International Conference on Communication Workshop, pp. 1161–1165 (2015)
5. Shahab, M.B., Kader, M.F., Shin, S.Y.: A virtual user pairing scheme to optimally utilize the spectrum of unpaired users in non-orthogonal multiple access. IEEE Signal Process. Lett. **23** (12), 1766–1770 (2016)

6. Wang, C.L., Chen, J.Y., Chen, Y.J.: Power allocation for a downlink non-orthogonal multiple access system. IEEE Wirel. Commun. Lett. **5**(5), 532–535 (2016)
7. Bi, Q., Liang, L., Yang, S.: Non-orthogonal multiple access technology for 5G systems **31** (5), 14–21 (2015)
8. Xu, P., Yuan, Y., Ding, Z., Dai, X., Schober, R.: On the outage performance of non-orthogonal multiple access with 1-bit feedback. IEEE Trans. Wirel. Commun. **15**(10), 6716–6730 (2016)
9. Dai, L., Wang, B., Yuan, Y., Han, S., Chin-Lin, I., Wang, Z.: Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends. IEEE Commun. Mag. **53**(9), 74–81 (2015)
10. Zhang, D.K.: Research on Power Allocation and Interference Elimination Algorithms for Non-Orthogonal Multiple Access System. Harbin Institute of Technology (2015)

Sum Secrecy Rate Maximization in NOMA Downlink SWIPT Systems

Jie Tang(✉), Tuwang Dai, Xiuyin Zhang, and Mo Huang

School of Electronic and Information Engineering,
South China University of Technology, Guangzhou, China
{eejtang,zhangxiuyin,mohuang}@scut.edu.cn

Abstract. In this paper, we investigate physical layer security in non-orthogonal multiple access (NOMA) downlink simultaneous wireless information and power transfer (SWIPT) systems, which consist of one access point (AP), information receivers (IRs), and energy receivers (ERs) that are capable of eavesdropping the IR. In order to achieve the target, maximizing the secure sum rate (SSR) of IRs under the conditions of satisfying both the IR's rate demand and the ER's harvested energy demand, an algorithm about power allocation of the AP is proposed. Due to the non-convexity of the SSR problem we simplify the objective function at first and decompose it into a series of subproblems, which are easily transformed into convex problems. Thus, the solution of subproblems can be derived according to the Karush–Kuhn–Tucker (KKT) conditions. Then, on account of the relationship between subproblems and original problem, the optimal power allocation is easily obtained. Using the numerical results, we demonstrate that the proposed algorithm can achieve better performance in NOMA downlink SWIPT system than the orthogonal multiple access (OMA) scheme respecting the SSR.

1 Introduction

On account of the wide application of intelligent devices and Internet of Things (IoT), the data traffic of the fifth generation (5G) is expected to increase by a factor of 1000 than now, and it is necessary that future 5G networks allow a large number of device access simultaneously [1]. In view of the above demands far surpassing the capabilities of the 4G's network, NOMA has been proposed and received widely in research since its advantageous in greatly improving the spectral efficiency of the system and accommodating massive connections [2].

Unlike the traditional OMA, NOMA enables to utilize the power domain for serving multiple users simultaneously with superposition coding, and it can significantly enhance the spectral efficiency of the system by sharing other orthogonal resources. Although it results in a large mutual interference while detecting the signal, the receiver can exactly decode the information by applying the successive interference cancellation (SIC) technology, and thus significantly improves the system performance [3]. In consequence, the combination of these two technologies can enhance the multi-user capacity [4].

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 790–802, 2019.
https://doi.org/10.1007/978-981-13-6264-4_93

In 5G networks, energy-constrained devices require longer operating times to provide better quality of service. In last few years, wireless power transfer (WPT) technology has made great progress, and it offers a new way to improve the operating time of energy-constrained wireless devices [5]. Motivated by the WPT, SWIPT has proposed lately intending to transmit both energy and information utilizing the same RF signal. Due to the great superiority of the two technologies, the combination of NOMA and SWIPT naturally attracted widespread attention and research for 5G networks [6].

However, wireless information is vulnerable to eavesdropping because of the broadcast characteristics of wireless transmission, it is an enormous risk for wireless transmission systems, especially the SWIPT system. Since the ERs require to be located closer to the AP than the IRs to meet harvested energy demands, thus they are surely capable of wiretapping the message transmitted to IRs. To meet this challenge, physical layer security technology is proposed as an alternative technology preventing information from wiretapping by utilizing the fading channels, and has been widely researched in wireless networks recently.

Different from the aforementioned literature, which only focus on investigating the system performance of SWIPT in association with NOMA or the physical layer security in conventional networks, we investigated the optimal power allocation in NOMA-downlink SWIPT system to optimize the SSR under the conditions of meeting the IR's rate demand and the ER's harvested energy demand. Because of the correlation between different coefficients, the optimization problem is obviously non-convex. To tackle this issue, we first simplify the initial problem by equivalent transformation and identify a feasible transmit power region, and then present an efficient power allocation policy and finally derive its closed-form expression on basis of KKT conditions. Numerical simulation results demonstrate that the algorithm achieves better system performance in NOMA downlink SWIPT system compared with other schemes, and the SSR performance of the system is analyzed with transmit power, the quality of service (QoS) demands, and the number of ERs, respectively.

2 System Model and Problem Formulation

We consider a downlink SWIPT enabled NOMA system, consisting of one base station (BS) with a single antenna, M single antenna IRs and K single antenna ERs. For each IR, all ERs may have better channel gain and are potential eavesdroppers, because the ERs are closer to the AP than IRs. Assuming that AP knows the perfect channel state information (CSI) of the IRs and ERs, which is practically valid because receivers are active in the system. In the process of signal transmission, the channel gain from the AP to the IRs consists of path loss and Rayleigh fading, and it can be denoted by $h_m = d_m^{-\frac{\alpha}{2}} v_m (m \in \{1, 2, ..., M\})$, where d_m denotes the distances between the AP and IRs, α denotes the path-loss exponent, v_m is the Rayleigh fading coefficients. Similarly, the channel gain of ERs is the same as IRs, denoted by $g_k = d_k^{-\frac{\alpha}{2}} v_k (k \in \{1, 2, ..., K\})$.

According to the NOMA protocol, the AP exploits the superposition coding and broadcasts the signal to all receivers. Therefore, the linear combined message can be formulated as $\sum_{m=1}^{M} \sqrt{a_m P} x_m$, where a_m denotes the power allocation coefficient constraint to $\sum_{m=1}^{M} a_m \leq 1$, P is the total transmit power of the AP, and x_m denotes the information symbol. Thus, the downlink signal received at the IRs is given by

$$y_m = h_m \sum_{m=1}^{M} \sqrt{a_m P} x_m + n_m, \forall m, \tag{1}$$

where n_m is the power of additive noise at IR. Similarly, the downlink received signal at the ERs, which might wiretap the message, is given by

$$y_k = g_k \sum_{m=1}^{M} \sqrt{a_m P} x_m + n_k, \forall k. \tag{2}$$

where n_k is the power of additive noise at IRs. Next, the harvested energy at the ERs can be given by

$$E_k^{ER} = \eta P |g_k|^2 \sum_{m=1}^{M} a_m, \forall k, \tag{3}$$

where $0 < \eta \leq 1$ denotes the energy conversion efficiency.

Similar to [7], we sort the channel coefficients as $0 < |h_1|^2 \leq |h_2|^2 \leq ... \leq |h_M|^2$ and $0 < |g_1|^2 \leq |g_2|^2 \leq ... \leq |g_K|^2$. For detecting their own information, all the IRs apply SIC to remove the mutual interference. SIC technology is based on the principle of gradually eliminating different user interference schemes. Specifically, data detection is performed one by one in order of transmission power, and then the interference is subtracted from the received signal, repeatedly until the own information is obtained. Thus, the achievable rate of IRs is given by

$$R_m^{IR} = \log_2 \left(1 + \frac{P|h_m|^2 a_m}{P|h_m|^2 \sum\limits_{i=m+1}^{M} a_i + \sigma^2} \right), \forall m. \tag{4}$$

In the meantime, the achievable rate at the ERs eavesdropping the information of the IRs are given by

$$R_{k,m}^{ER} = \log_2 \left(1 + \frac{P|g_k|^2 a_m}{P|g_k|^2 \sum\limits_{i=m+1}^{M} a_i + \sigma^2} \right), \forall k, m. \tag{5}$$

We assume that the potential eavesdropper in the ERs has successfully decoded the information of the first $m-1$ IRs while trying to wiretap to the m-th. Hence, in the case of the ERs acting as eavesdroppers, the achievable secure

rate at the IRs is given by

$$R_{k,m}^{s} = \left[R_{m}^{IR} - R_{k,m}^{ER}\right]^{+}, \tag{6}$$

$$R_{s} = \sum_{m=1}^{M} \min_{k} R_{k,m}^{s}, \tag{7}$$

where $[y]^{+} \triangleq \max(0, y)$.

Therefore, the corresponding SSR optimization problem can be mathematically written as

$$\max_{a_m, 1 \le m \le M} R_{s} = \sum_{m=1}^{M} \min_{k} R_{k,m}^{s} \tag{8a}$$

$$\text{s.t.} \quad R_{m}^{IR} \ge Q_{m}, \forall m \tag{8b}$$

$$E_{k}^{ER} \ge E_{k}, \forall k \tag{8c}$$

$$\sum_{m=1}^{M} a_{m} \le 1, \tag{8d}$$

where Q_m and E_k denote the QoS demand and the harvested energy demand, respectively.

In consideration of the SSR optimization problem as non-convex, the optimal power allocation can be derived by an exhaustive search method with unacceptable complexity. To reduce complexity, we will propose an effective algorithm to deal with this problem.

3 Power Allocation Algorithm

In this section, an effective algorithm is presented to work out the problem of maximizing the SSR in the NOMA downlink SWIPT system. We simplify the objective function at first and determine the minimum transmit power. Next, decomposing the initial problem into a series of subproblems, and the solution of subproblems can be derived according to the KKT conditions. Finally, on account of the relationship between subproblems and original problem, the optimal power allocation is easily obtained.

3.1 Problem Simplification

According to constraint condition (8b) and (8c), these constraints are reformulated as follows:

$$a_{m} \ge A_{m} \left(P|h_{m}|^{2} \sum_{i=m+1}^{M} a_{i} + \sigma^{2} \right), \forall m, \tag{9}$$

$$P\sum_{m=1}^{M} a_m \geq \frac{E_k^{ER}}{\eta|g_k|^2}, \forall k, \tag{10}$$

where $A_m = \dfrac{2^{Q_m}-1}{P|h_m|^2}$.

To simplify the SSR optimization problem, we simplify the objective function at first according to the theorem as follows.

Theorem 1. *For all IRs, the better the channel of ER is, the lower the SSR is, i.e.,*

$$\min_{1\leq k\leq K} R_{k,m}^s = R_{K,m}^s, \forall m.$$

Proof. For simplicity, we perform variable substitution on formula (5), as follows:

$$f(x) = \log_2\left(1+\frac{Ax}{Bx+\sigma^2}\right), \tag{11}$$

where $A = Pa_m$, $B = P\sum_{i=m+1}^{M} a_i$, and $x = |g_k|^2$. Therefore, the first-order derivative of function $f(x)$ on x is obtained, as shown below:

$$\frac{df(x)}{dx} = \frac{A\sigma^2}{\ln 2\left(1+\dfrac{Ax}{Bx+\sigma^2}\right)(Bx+\sigma^2)^2}. \tag{12}$$

Obviously, $f(x)$ is a monotonically increasing function of x for $\dfrac{df(x)}{dx} > 0$. It indicates that as the channel gain $|g_k|^2$ improves, the $R_{k,m}^{ER}$ increases. Thus, we can get the conclusion: $\min_k R_{k,m}^s = R_{K,m}^s$. The proof is completed. ■

Based on the above conclusion, we resort the channel gain of the IRs and the K-th ER, which have the best channel gain among the ERs, with $0 < |h_1|^2 \leq |h_2|^2 \leq ... \leq |h_{M_k}|^2 \leq |g_K|^2 \leq |h_{M_k+1}|^2 \leq ... \leq |h_M|^2$. Hence, the SSR of the system, denoted by R_s, is given by

$$R_s = \sum_{m=M_k+1}^{M} \left(R_m^{ID} - R_{K,m}^{ER}\right). \tag{13}$$

Thus, the initial SSR problem in (8) is reformulated as

$$\max_{a_m, 1\leq m\leq M} R_s = \sum_{m=M_k+1}^{M} \left(R_m^{ID} - R_{K,m}^{ER}\right) \tag{14a}$$

$$\text{s.t.} \quad a_m \geq A_m\left(P|h_m|^2\sum_{i=m+1}^{M} a_i + \sigma^2\right), \forall m \tag{14b}$$

$$P\sum_{m=1}^{M} a_m \geq \frac{E_k^{ER}}{\eta|g_k|^2}, \forall k \tag{14c}$$

$$\sum_{m=1}^{M} a_m \leq 1. \tag{14d}$$

3.2 Minimum Transmit Power

Observing the constraint conditions (14b) and (14c), the transmit power of AP should be not less than a certain transmit power that satisfies both the demands of IRs and ERs. Thus, we first determine the minimum transmit power of the AP, before solving the SSR problem (14).

From constrain conditions (14b), the transmit power P should be equal or greater than a certain transmit power P_{min_ir}, to meet the IR's QoS demands. This certain transmit power is derived when the constrain conditions (8b) are equal, i.e., $R_m^{IR} = Q_m, \forall m$ [7]. Thus, P_{min_ir} can be given by

$$P_{min\_ir} = \sum_{m=1}^{M} a_m^{min} P, \tag{15}$$

where $a_m^{min} (m \in \{1, 2, ..., M\})$ denotes the minimum power allocation coefficient just to meet the IR's rate demand. It should be noted that $a_m^{min} (m \in \{1, 2, ..., M\})$ can be determined sequentially in the order of $M, M-1, ..., 1$ when the inequality in (14b) is set to equality.

Similarly, in order to demand the ERs' harvested energy demand, there is existing another certain transmit power at the AP, denoted by P_{min_er}. And, it is easy to get this power P_{min_er} only when the inequality in (8c) is active, i.e., $E_k^{ER} = E_k, \forall m$. Therefore, on basis of (8d) and (10), we can derive the minimum transmit power P_{min_er} as

$$P_{min\_er} = \max_k \left\{ \frac{E_k^{ER}}{\eta |g_k|^2} \right\}. \tag{16}$$

In brief, if and only if the P is larger than the P_{min}, which just meets the QoS of IRs and ERs, the SSR problem has a feasible solution. The P_{min} is derived as follows:

$$P \geq P_{min} = \max\{P_{min\_ir}, P_{min\_er}\}, \tag{17}$$

and it can be assumed that the achieved SSR of the NOMA downlink SWIPT system is zero, when the AP transmit power $P < P_{min}$.

3.3 Optimal Power Allocation Strategy

According to the above steps, we can obtain the P_{min}. Therefore, we will address the SSR optimization problem under given P_{min}. Hence, optimization problem (14) is reformulated as

$$\max_{a_m, 1 \leq m \leq M} R_s = \sum_{m=M_k+1}^{M} \left(R_m^{ID} - R_{K,m}^{ER} \right) \tag{18a}$$

$$\text{s.t.} \quad a_m \geq A_m \left(P |h_m|^2 \sum_{i=m+1}^{M} a_i + \sigma^2 \right), \forall m \tag{18b}$$

$$\sum_{m=1}^{M} a_m \leq 1. \tag{18c}$$

Next, substituting formula (4) and formula (5) into formula (13) at first, and the objective function of (18) can be rewritten as

$$\begin{aligned} R_s = & \sum_{m=M_k+1}^{M} \left[\log_2 \left(1 + \frac{P|h_m|^2 a_m}{P|h_m|^2 \sum_{i=m+1}^{M} a_i + \sigma^2} \right) \right. \\ & \left. - \log_2 \left(1 + \frac{P|g_K|^2 a_m}{P|g_K|^2 \sum_{i=m+1}^{M} a_i + \sigma^2} \right) \right] \\ = & \log_2 \left(P|h_{M_k+1}|^2 \sum_{i=M_k+1}^{M} a_i + \sigma^2 \right) \\ & - \log_2 \left(P|g_K|^2 \sum_{i=M_k+1}^{M} a_i + \sigma^2 \right) \\ & + \sum_{m=M_k+1}^{M-1} \left[\log_2 \left(P|h_{m+1}|^2 \sum_{i=m+1}^{M} a_i + \sigma^2 \right) \right. \\ & \left. - \log_2 \left(P|h_m|^2 \sum_{i=m+1}^{M} a_i + \sigma^2 \right) \right]. \end{aligned} \tag{19}$$

To simplify Eq. (19), we define

$$C_m \triangleq \begin{cases} P|g_K|^2, m = M_k \\ P|h_m|^2, M_k + 1 \le m \le M \end{cases}$$

$$t_m \triangleq \sum_{i=m+1}^{M} a_i, M_k \le m \le M - 1$$

and

$$f_m(t_m) \triangleq \log_2 \left(C_{m+1} t_m + \sigma^2 \right) - \log_2 \left(C_m t_m + \sigma^2 \right).$$

Then, (19) can be rewritten as

$$R_s = \sum_{m=M_k}^{M-1} f_m(t_m). \tag{20}$$

From (20), it is obviously the simplified objective function R_s which is still non-convex, and R_s can be considered as the sum of $M - M_k$ sub-functions. Thus, we can decompose the problem (18) into $M - M_k$ subproblems, where these subproblems are to maximize $f_m(t_m)(M_k \le m \le M)$ under the constraints on (18), respectively. Then, on account of the relationship between subproblems and original problem, the optimal policy is easily determined. In the following, we provide detailed procedures for solving the subproblems.

To solve the subproblems, we can utilize the monotonicity of the function $f_m(t_m)$ and its first-order derivative on t_m is given by

$$\frac{df_m(t_m)}{dt_m} = \frac{(C_{m+1} - C_m)\sigma^2}{\ln 2(C_{m+1}t_m + \sigma^2)(C_m t_m + \sigma^2)}. \tag{21}$$

From (21), we can derive that $\frac{df_m(t_m)}{dt_m} \geq 0$ because $C_{m+1} > C_m$. Therefore, $f_m(t_m)$ is a monotonically increasing function of t_m. Therefore, the subproblem for maximizing the $f_m(t_m)$ is given by

$$\max_{a_m, 1 \leq m \leq M} \; t_m \tag{22a}$$

$$\text{s.t.} \quad a_m \geq A_m \left(P|h_m|^2 \sum_{j=m+1}^{M} a_j + \sigma^2 \right), \forall m \tag{22b}$$

$$\sum_{m=1}^{M} a_m \leq 1. \tag{22c}$$

Apparently, the problem (22) is convex. Thus, the KKT conditions of the problem (22) can be easily derived as follows:

$$\lambda = \begin{cases} \mu_k - \sum_{i=1}^{k-1} \mu_i A_i P|h_i|^2, & 1 \leq k \leq m \\ \mu_k - \sum_{i=1}^{k-1} \mu_i A_i P|h_i|^2 + 1, & m < k \leq M, \end{cases} \tag{23}$$

$$\mu_i \left[A_i \left(P|h_i|^2 \sum_{j=i+1}^{M} a_j + \sigma^2 \right) - a_i \right] = 0, 1 \leq i \leq M \tag{24}$$

$$\mu_i \geq 0, \; 1 \leq i \leq M \tag{25}$$

$$\lambda \left(\sum_{i=1}^{M} a_i - 1 \right) = 0 \tag{26}$$

$$\lambda \geq 0, \tag{27}$$

where $\{\mu_i\}_{i=1}^{M}$ and λ denote the Lagrange factors for the constraint condition (22b) and (22c), respectively.

Inspired by the results in [7], and by analyzing above KKT conditions, the closed-form solution of the problem (22) is given by

$$a_i = \frac{A_i \left[P|h_i|^2 \left(1 - \sum_{j=1}^{i-1} a_j \right) + \sigma^2 \right]}{2^{Q_i}}, 1 \leq i \leq m, \tag{28}$$

$$t_m = 1 - \sum_{i=1}^{m} a_i. \tag{29}$$

Based on the solution of subproblem, we can derive that when $m = M-1$, the solution of the subproblem is the solution of the initial problem (22a), because the solution satisfies all the constrains of all subproblems, so the optimal power allocation coefficients can be given by

$$a_m^\star = \begin{cases} \dfrac{\left(2^{Q_m}-1\right)\left[P|h_m|^2\left(1-\sum_{i=1}^{m-1} a_i^\star\right)+\sigma^2\right]}{P|h_m|^2 2^{Q_m}}, & 1 \le m \le M-1, \\ 1-\sum_{i=1}^{M-1} a_i^\star, & m = M. \end{cases} \tag{30}$$

4 Simulation Results and Discussion

In this section, we evaluate the SSR performance of NOMA downlink SWIPT system with the proposed optimal policy via numerical simulations. In the simulation setup, we assumed that the distance between AP and receivers is 80 m, the path-loss exponent $\alpha = 2.5$ [8], and the energy conversion efficiency $\eta = 10\%$ [9]. In order to simplify the calculation, it can be assumed that all IRs have the

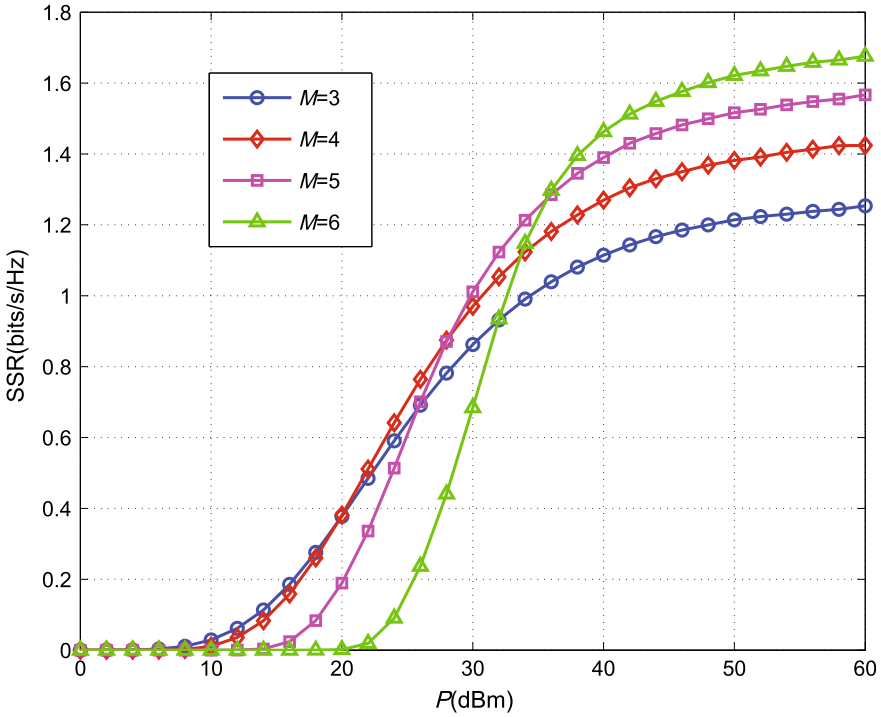


Fig. 1. The SSR performance in terms of the total transmit power of the AP P with $\bar{Q} = 2.5$ bit/s/Hz, $\bar{E} = -50$ dBm, and K=2

same QoS demand and all ERs have the same harvested energy demand, i.e., $Q_m = \bar{Q}$, $1 \leq m \leq M$, $E_k = \bar{E}$, $1 \leq k \leq K$. What is more, when P is not large enough to meet all IRs data rate demands and all ERs harvested energy demands, that is, $P < P_{min}$, R_s is set to zero.

Figure 1 shows the SSR performance of NOMA downlink SWIPT system with different numbers of receivers as a function of the total transmit power P. As it can be observed from Fig. 1, the achievable SSR by the presented optimal policy is monotonically nondecreasing versus the transmit power P. Actually, the transmit power P is lower than the minimum power, i.e., $P < 8$ dbm, the sum secrecy rate is zero because P is too small to satisfy all IRs minimum rate demands and all ERs harvested energy demands. What is more, the SSR performance shows an upward trend as the number of IRs increases when $P > 36$ dbm. This is because, if the M is larger, the diversity gain is higher, though it needs more power to meet the QoS demands.

Next, it is investigated that the influence of the QoS demand on the achieved SSR performance, shown in Figs. 2 and 3. As a result, Fig. 2 shows that when $\bar{Q}$ increases, the achievable SSR decreases. This is because, with $\bar{Q}$ increasing, the AP requires to use more power to satisfy the Qos demand of IRs, which have

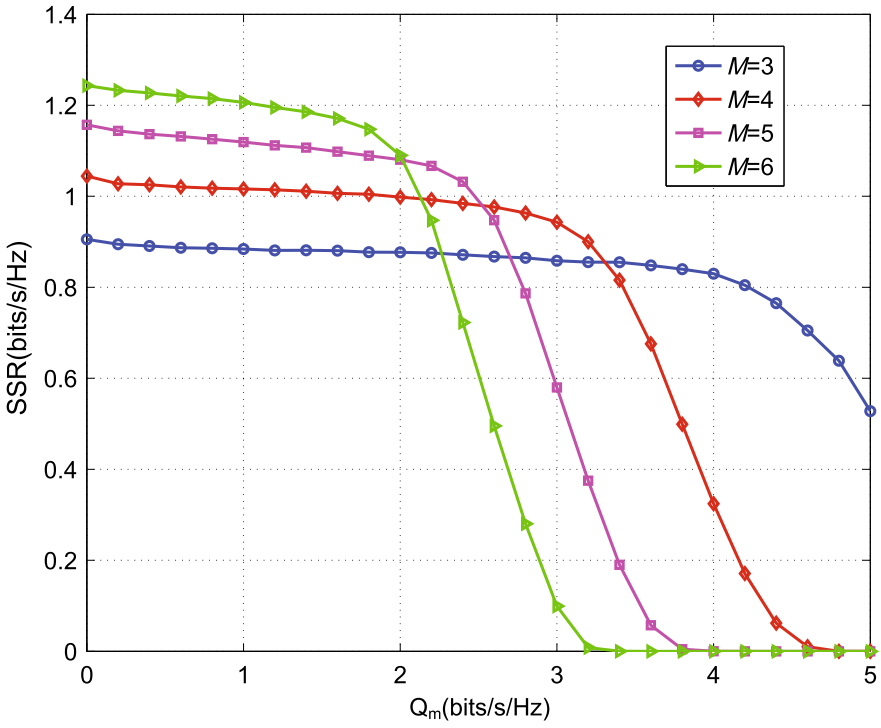


Fig. 2. The SSR performance in terms of the rate demand $\bar{Q}$, i.e., Q_m, for different numbers of IRs with $P = 30$ dBm, $\bar{E} = -50$ dBm, and $K = 2$

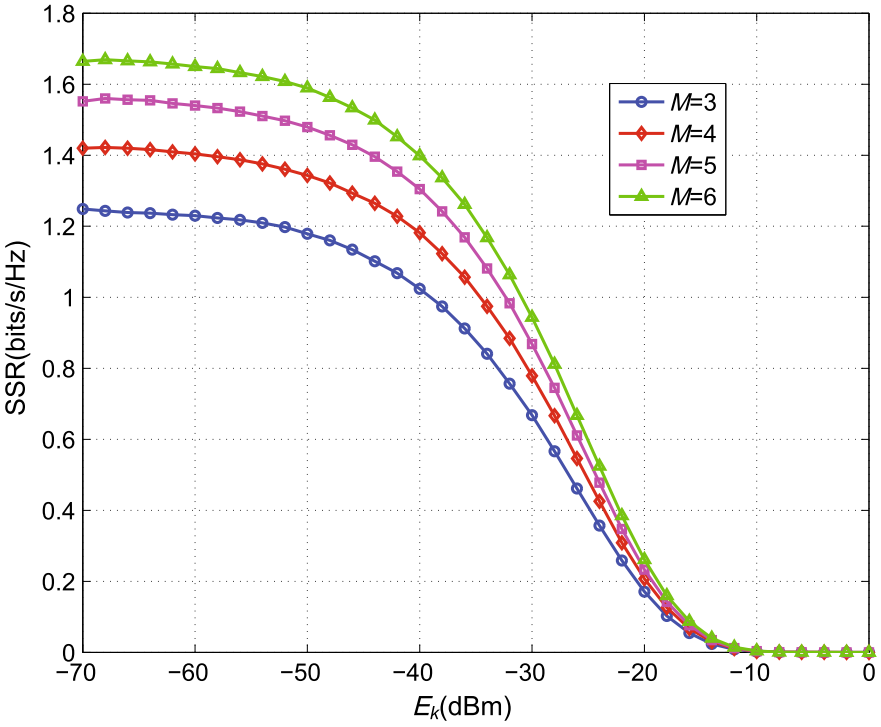


Fig. 3. The SSR performance in terms of the harvested energy demand $\bar{E}$, i.e., E_k with $P = 30\,\text{dBm}$, $\bar{Q} = 2.5$ bit/s/Hz, and $K = 2$

the poor channel condition. Apparently, it will degrade the system performance. Furthermore, if $\bar{Q}$ is too high, the achievable SSR is zero, which indicates that the AP total transmit power is not large enough to meet all IRs Qos demands. Besides, as the number of IRs decreases, the achieved SSR more slowly declines. Then, the influence of the harvested energy demand $\bar{E}$ on the SSR performance is shown in Fig. 3. As expected, the SSR performance shows a downward trend as $\bar{E}$ increases. In addition, the achievable SSR will be zero as $\bar{E}$ becomes larger since the AP transmit power is too small to meet the harvested energy demands of all ERs. Thus, the system parameter should be selected based on a trade-off between receivers' demand and the SSR performance.

To further evaluate the superiority of the proposed policy in NOMA downlink SWIPT scheme (denoted by "SWIPT+NOMA"), in this simulation, two different downlink schemes are presented as benchmark. In first benchmark scheme, we consider the traditional OMA scheme that the IRs are allocated the same spectrum resources and transmitted power. In the second benchmark scheme, SWIPT combines with the traditional OMA is considered (denoted by "SWIPT+OMA"). According to Fig. 4, it can be obviously seen when P is larger than 28 dBm, the SSR performance of the presented scheme is far better than two

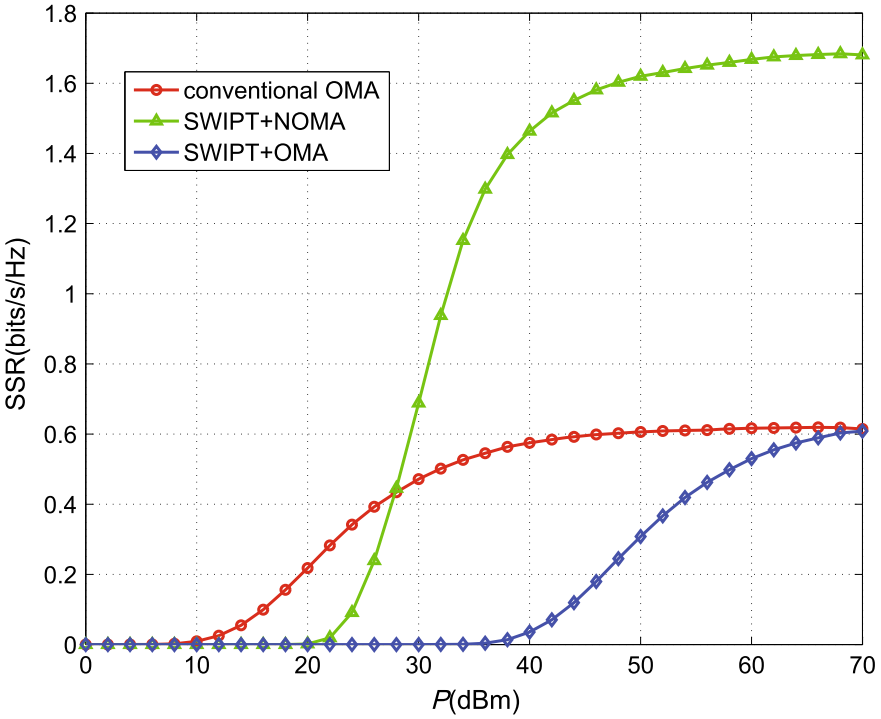


Fig. 4. The SSR performance of different schemes in terms of the total transmit power P with $\bar{Q} = 2.5\,\text{bit/s/Hz}$ and $\bar{E} = -50$ dBm

benchmark schemes. However, P is lower than a certain power, i.e., $P < 28\,\text{dbm}$, the traditional OMA scheme has the best SSR performance. The reason is that the traditional OMA scheme does not consider the harvested energy demands of ERs, just have to satisfy the rate demands of all IRs.

5 Conclusion

In this paper, we considered physical layer security in a NOMA downlink SWIPT SISO systems consisting of multiple IRs and multiple ERs, which likely to eavesdrop the information. Our objective is to optimize the SSR under the constraints on the QoS demand and the harvested energy demand by optimizing the power allocation of the AP. To address the non-convex optimization problem, we first find the feasible power region, and then decompose the original problem into a series of subproblems since the solution of subproblems can be derived according to the KKT conditions. Then, based on the relationship of subproblems and original problem, we can easily derive the optimal power allocation. Simulation results have proven the effectiveness of the proposed algorithm to the SSR maximization. More significantly, our proposed algorithm in the NOMA downlink SWIPT system has better SSR performance than the traditional OMA approach.

References

1. Li, Q.C., Niu, H., Papathanassiou, A.T., Wu, G.: 5G network capacity: key elements and technologies. IEEE Veh. Technol. Mag. **9**, 71–78 (2014). Mar.
2. Boccardi, F., Heath, R.W., Lozano, A., Marzetta, T.L., Popovski, P.: Five disruptive technology directions for 5G. IEEE Commun. Mag. **52**, 74–80 (2014). Feb
3. Dai, L., Wang, B., Yuan, Y., Han, S., Chih-Lin., I., Wang, Z.: Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends. IEEE Commun. Mag. **53**, 74–81 (2015)
4. Li, A., Lan, Y., Chen, X., Jiang, H.: Non-orthogonal multiple access (NOMA) for future downlink radio access of 5G. China Commun. **12**, 28–37 (2015). Dec.
5. Lu, X., Wang, P., Niyato, D., Kim, D.I., Han, Z.: Wireless networks with RF energy harvesting: a contemporary survey. IEEE Commun. Surv. Tuts. **17**, 757–789 (2015). (Secondquarter)
6. Yang, Z., Ding, Z., Fan, P., Al-Dhahir, N.: The impact of power allocation on cooperative non-orthogonal multiple access networks with SWIPT. IEEE Trans. Wirel. Commun. **16**, 4332–4343 (2017). July
7. Zhang, Y., Wang, H.M., Yang, Q., Ding, Z.: Secrecy sum rate maximization in non-orthogonal multiple access. IEEE Commun. Lett. **20**, 930–933 (2016). May
8. IEEE P802.11 Wireless LANs, TGn Channel Models, IEEE 802.11-03/940r4, Technical Report, May 2004
9. Krikidis, I., Timotheou, S., Nikolaou, S., Zheng, G., Ng, D.W.K., Schober, R.: Simultaneous wireless information and power transfer in modern communication systems. IEEE Commun. Mag. **52**(11), 104–110 (2014). Nov

An Improved Detector Design for Combining Power Domain NOMA to Spectrally Efficient FDM

Min Jia[1,2(✉)], Qiling Gao[1], Zhisheng Yin[1], Qing Guo[1], and Xuemai Gu[1]

[1] School of Electronics and Communication Engineering, Communications Research Center, Harbin Institute of Technology, Harbin, China
jiamin@hit.edu.cn
[2] CETC Key Laboratory of Aerospace Information Applications, Beijing, China

Abstract. In this paper, the system combining Non-Orthogonal Multiple Access (NOMA) with Spectrally Efficient Frequency Division Multiplexing (SEFDM) is mainly discussed. First, the mathematical model of NOMA–SEFDM system is given to further improve bandwidth efficiency. The widely used structure of Successive Interference Cancellation (SIC) is succeeded and the performance of NOMA–SEFDM system is investigated with Maximum Likelihood (ML) detector. Moreover, for BPSK modulation, considering the two users' transmit symbols as a four-point constellation, the feasibility of iterative detection in NOMA–SEFDM system is verified, which achieves a quasi-optimal performance while the number of iterations is moderate while behaves with low computing complexity.

Keywords: NOMA–SEFDM · SIC power domain multi-users

1 Introduction

Orthogonal Frequency Division Multiplexing (OFDM) technology is interested as a solution for providing high-rate data transmission service [1]. In OFDM system, the bandwidth is divided into multiple parallel orthogonal subcarriers and the interval between adjacent subcarriers satisfies orthogonality. Even the subcarriers have 50% overlapping, OFDM system can achieve a higher spectrum efficiency compared to traditional Frequency Division Multiple Access (FDMA) [2]. For 4G standard, OFDM is a technology used in physical layer; however, it is sensitive to frequency offset, which leads to Inter-carrier Interference (ICI) [3]. With the demand for high-speed communication and high-efficiency spectrum utilization, the type of technology using non-orthogonal multicarrier has raised, such as Spectrally Efficient Frequency Division Multiplexing (SEFDM). Early application of this kind of system can be tracked back to 2003 [4, 5], the main idea of SEFDM system is through compressing the space between adjacent subcarrier to improve spectral efficiency. As the previous research could see, while bandwidth compression factors up to 25%, it can achieve the same performance as OFDM system [6]. By the same bandwidth, compared with OFDM, SEFDM offers a

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 803–810, 2019.
https://doi.org/10.1007/978-981-13-6264-4_94

higher data transmission speed. In other words, SEFDM requires less bandwidth for the same data speed [7, 8]. Also, the key problem is ICI caused by compressing subcarrier spacing. Therefore, additional detectors are required at receiver to detect signals which are disturbed by ICI due to the loss of orthogonality [9]. Iterative detection algorithm has been proposed to solve this problem in SEFDM system but it is not very efficient for high bandwidth compression [10]. While another algorithm maximum likelihood is widely used because of the optimal Bit Error Rate (BER) performance, inevitably it has the weakness of high computational complexity. It is worth noting that such detectors are all limited to small system size because complexity is getting higher as the number of subcarriers increase [11, 12].

As another non-orthogonal technology to further enhance the spectral efficiency, Non-orthogonal Multiple Access (NOMA) has been considered in this work. NOMA system is a power domain resource allocation scheme [13], and for the sake of multiple access, many users occupy the same time, frequency, and code resource but different power factor to distinguish users. There is no doubt that the most prominent advantage is to improve spectrum efficiency, besides, a more flexible and high-efficient user access is allowed in NOMA system [14–16]. And it is shown that NOMA can achieve a higher throughput compared with orthogonal multiple access [17]. Obviously, power domain reuse will introduce interference between users, normally, NOMA based on SIC algorithm is designed to enhance the performance. The main principle of SIC is to detect users in the order of Signal-to-Noise Ratio (SNR), the performances of users are positively related to the sequence of detection [18].

In this paper, the multiple access and transmission design are all presented as non-orthogonal characteristic through combining NOMA with SEFDM, which is expected to provide much higher spectrum efficiency and faster information rate, also allows more flexible user access. As the idea of the proposed structure is carried out, the corresponding receiver design and signal demodulation have to be updated to a new revision. Focusing on this issue, SIC algorithm is investigated first for power domain multiple users multiplexing and the classical ML and ID are also introduced to the proposed NOMA–SEFDM system.

2 System Model

2.1 NOMA–SEFDM Transmitter

The combination of NOMA system and SEFDM system is mainly used to get high spectrum efficiency and a flexible user access. The sententious block diagram of the transmitter of SEFDM system is shown in Fig. 1. In reality, non-orthogonal subcarriers are used to transmit symbols and it is worth noting that the symbols are the summation of multi-users' signal in power domain.

When the account of users enlarged to m, the expression of transmitted signal is

$$S = \sum_{i=1}^{m} \sqrt{p_i} s_i, \tag{1}$$

while in (1), p_i is the power factor applied to users, there should be $\sum_{i=1}^{m} p_i = 1$ and s_i denotes the modulated symbols of user i.

Then, the summation signals of m users are transferred by N subcarriers which are completely non-orthogonal.

In a normal OFDM system, in order to satisfy the orthogonality, the minimal distance between adjacent subcarrier is defined as $\Delta f = \frac{1}{T}$, while for SEFDM system, through compress the distance between subcarrier defined by OFDM to get a higher spectral efficiency, the frequency spacing can be defined as $\Delta f = \frac{\alpha}{T}$, where α is bandwidth compression factor and $0 < \alpha < 1$. Transmitted signal in NOMA–SEFDM system is expressed as

$$x(t) = \frac{1}{\sqrt{T}} \sum_{l=-\infty}^{+\infty} \sum_{n=0}^{N-1} S_{l,n} \exp\left[\frac{j2\pi n\alpha(t - lT)}{T}\right], \tag{2}$$

where N represents the number of subcarriers and $S_{l,n}$ is the l^{th} symbol modulated on n^{th} subcarrier. T is the period of one SEFDM symbol. The percentage of bandwidth saving equals to $(1 - \alpha) \times 100\,\%$.

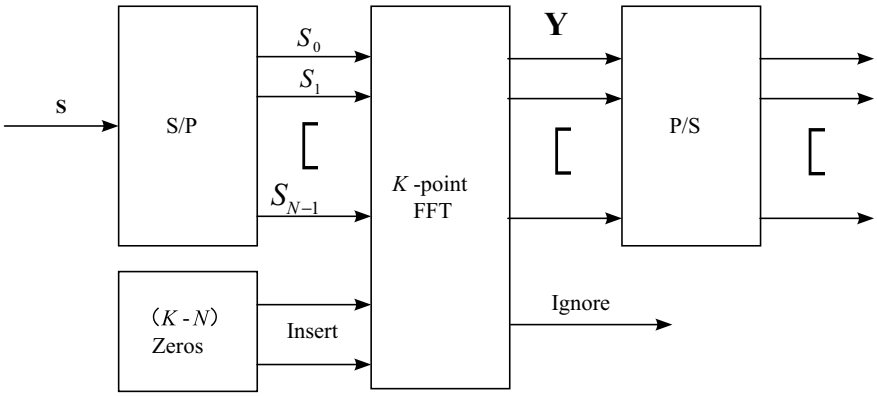


Fig. 1. Block diagram of SEFDM transmitter

2.2 NOMA–SEFDM Receiver

We give a block diagram of NOMA–SEFDM receiver corresponding to transmitter in Fig. 2. Presume that the SEFDM signal in Eq. (2) passing through additive white Gaussian noise channel, and at receiver it can be defined as

$$y(t) = x(t) + n(t). \tag{3}$$

While after fast Fourier transform, the signal can be simply expressed in matrix as

$$\mathbf{Y} = \mathbf{C}\mathbf{y} + \mathbf{N}, \tag{4}$$

where $\mathbf{Y}$ is an N-dimensional vector of transmitted signals after a Fast Fourier Transform (FFT) operation, $\mathbf{C}$ is an $N \times N$ correlation matrix that describes the interference between subcarriers. Where $\mathbf{C}$ is defined as $\mathbf{C} = \mathbf{F}^*\mathbf{F}$, $\mathbf{F}$ denotes the subcarrier matrix with elements equal to $e^{\frac{j2\pi nk\alpha}{N}}$, $\mathbf{F}^*$ is the conjugate subcarrier matrix with elements equal to $e^{\frac{-j2\pi nk\alpha}{N}}$ and N represents the N-dimensional vector of AWGN.

After SEFDM detector such as ML detector, the standard modulated constellation is recovered, where R denotes multi-user's sum signal.

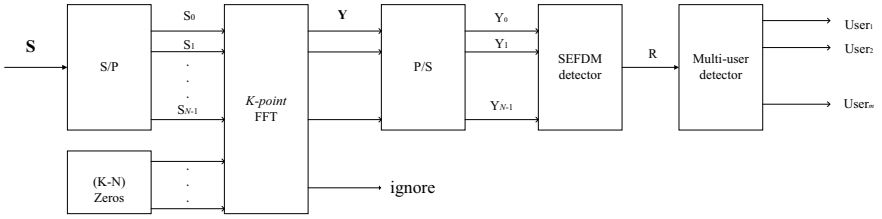


Fig. 2. Block diagram of NOMA–SEFDM receiver

3 Receive Detector

For traditional iterative detection under single user SEFDM system, the pattern of modulation decides the number of constellation points. In this paper, an iterative detection is proposed to demodulate multi-user information. The main idea of this algorithm is defined by

$$S_n = \lambda\mathbf{R} + (e - \lambda\mathbf{C})S_{n-1}, \tag{6}$$

where S_n is recovering symbols after n iteration, S_{n-1} is an N-dimensional vector of estimated symbols after n iterations, e is identity matrix, and λ is convergence factor which is set to 1 in this paper. Figure 3 shows the points of summation signals and we have defined the uncertain interval by variate $d = \sqrt{p_2}(1 - m/v)$, where m is the m_{th} iteration and v is total number of iterations.

Only when points fall in the allowable area can be mapped to corresponding constellation while others are keeping unchanged to next iteration. In accordance with the principle of iteration detection, uncertainty area is reduced after each iteration. It should be noted that for multi-users, we assume that the transmit signal is approximate to constellations, iteration detection is appropriate to reduce complexity while the performance is close to maximum likelihood.

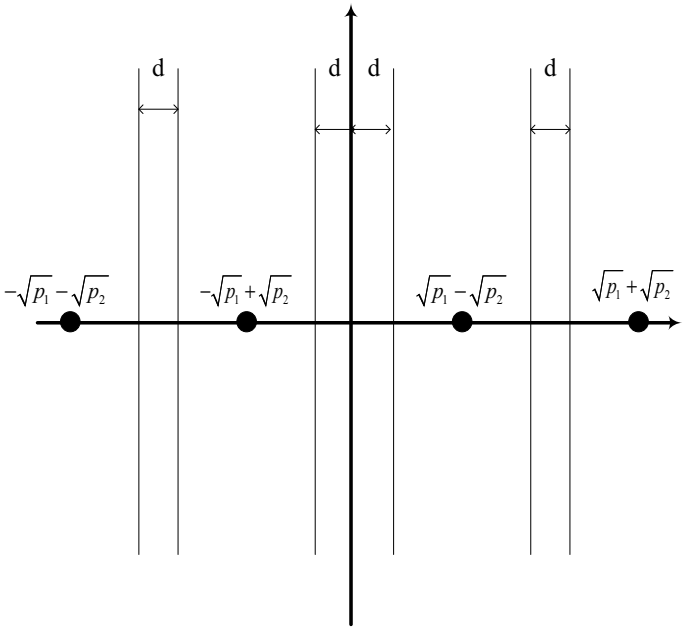


Fig. 3. Soft mapping of summation signal under BPSK modulate

The information of each user is exactly recovered from signal S after ML detector. In Fig. 4, we describe a block diagram about SIC detector, according to SNR in receiver to demodulate user's information, the highest power user is demodulated first and subtracted, then demodulate and subtract the second highest user, until all users' information are recovered completely.

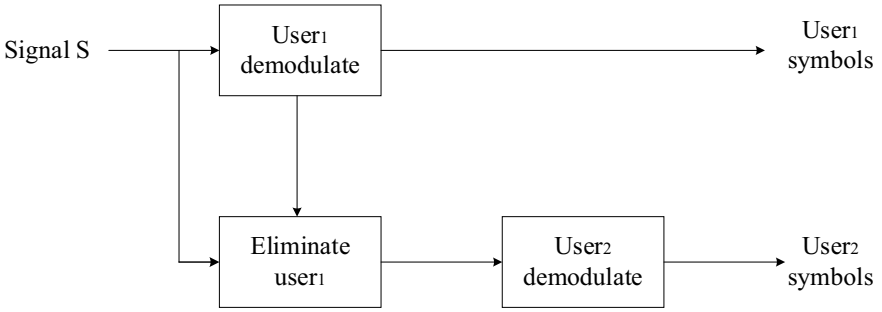


Fig. 4. Block diagram of SIC detector of two users

The details of the ID-SIC detector are described in Algorithm 1. $\hat{\mathbf{S}}$ is the expression of unconstrained estimated symbols and $\tilde{\mathbf{S}}$ defines an N-dimensional vector of constrained estimated symbols. λ is a convergence factor between 1 and 2, v is total

number of iterations, and m is m_{th} iteration.

Algorithm1: ID-SIC algorithm

Input: $\mathbf{R}$, $\mathbf{C}$, v, λ ;

Output: users' signal $\mathbf{X}_1, \mathbf{X}_2$;

Initialization:

$\tilde{\mathbf{S}} = \mathbf{R}$;

Repeat:

1. $\hat{\mathbf{S}} = \lambda R + (e - \lambda \mathbf{C})\tilde{\mathbf{S}}$;
2. $d = \sqrt{p_2}(1 - m/v)$;
3. mapping signal $\hat{\mathbf{S}}$,if belongs to allowable region, sentence $\tilde{\mathbf{S}}$ as a certain point, else $\tilde{\mathbf{S}} = \hat{\mathbf{S}}$;

Until $m == v$;

$\mathbf{S} = \tilde{\mathbf{S}}$;

$\mathbf{X}_1 = Demo(\mathbf{S})$;

$\mathbf{X}_2 = \mathbf{S} - Mod(\mathbf{X}_1)$;

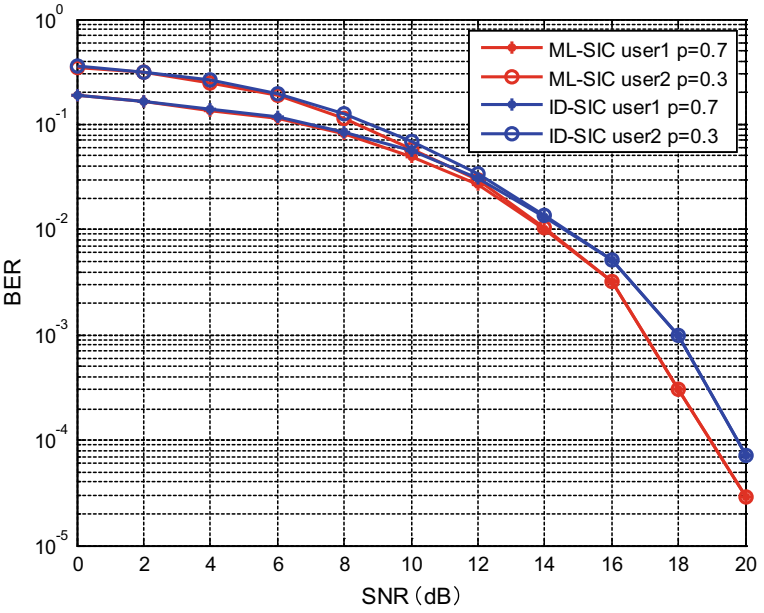


Fig. 5. BER performance for ID and ML detectors carrying BPSK symbols for number of iterations is 300, power allocation factor

4 Numerical Results

The performance and complexity of the proposed detector and algorithm are evaluated. The work in this paper is to prove a concept of NOMA with non-orthogonal subcarrier transmission mode. Therefore, the modulation scheme adopted in this simulation is BPSK, the number of users is set to 2 and only an AWGN channel is assumed.

Performance for ID-SIC detector and ML-SIC detector is shown in Fig. 5, for the same number of subcarriers, ID-SIC detector has an approximate behavior compared with ML-SIC detector when the number of iterations has been set to 300, and the complexity of ID-SIC detector is lower than ML-SIC detector according to the principle mentioned before.

5 Conclusion

In this paper, we proposed a system combining NOMA with SEFDM transmission system, with superiorities such as high spectral efficiency and high user capacity, sets in both power domain multiplexing and bandwidth saving. Accordingly, receiver design such as ML-SIC detector which is adapted to NOMA–SEFDM system has been raised in order to solve the problems above. Moreover, the feasibility of iteration detection for summation symbols of multi-users is verified, which could achieve a lower complexity while maintaining a quasi-optimal performance compared with ML detector at a proper number of iterations.

Acknowledgements. This work was supported by the National Natural Science Foundation of China under Grant 61671183, Grant 61771163, and the Open Research Fund of CETC key laboratory of aerospace information applications under Grant SXX18629T022.

References

1. Sahin, A., Guvenc, I., Arslan, H.: A survey on multi carrier communications: prototype filters, lattice structures, and implementation aspects. IEEE Commun. Surv. Tutor. **16**(3), 1312–1338 (2014)
2. Xu, T., Darwazeh, I.: Multi-sphere decoding of block segmented SEFDM signals with large number of sub-carriers and high modulation order. In: 2017 International Conference on Wireless Networks and Mobile Communications (WINCOM), Rabat, pp. 1–6 (2017)
3. Rodrigues, M., Darwazeh, I.: A spectrally efficient frequency division multiplexing based communications system. In: Proceedings of 8th International OFDM Workshop, Hamburg, pp. 48–49 (2003)
4. Xu, T., Darwazeh, I.: M-QAM signal detection for a non-orthogonal system using an improved fixed sphere decoder. In: 2014 9th International Symposium on Communication Systems, Networks & Digital Sign (CSNDSP), Manchester, pp. 623–627 (2014)
5. Ghannam, H., Darwazeh, I.: SEFDM over satellite systems with advanced interference cancellation. IET Commun. **12**(1), 59–66 (2018)
6. Wang, Y., et al.: SEFDM based spectrum compressed VLC system using RLS time-domain channel estimation and ID-FSD hybrid decoder. In: 42nd European Conference on Optical Communication, ECOC 2016, Dusseldorf, Germany, pp. 1–3 (2016)

7. Luo, F-L., Zhang, C.: Bandwidth-compressed multicarrier communication: SEFDM. In: Signal Processing for 5G: Algorithms and Implementations, vol. 1, p. 616. Wiley-IEEE Press (2016)
8. Xu, T., Darwazeh, I.: A soft detector for spectrally efficient systems with non-orthogonal overlapped sub-carriers. Commun. Lett. IEEE **18**(10), 1847–1850 (2014)
9. Darwazeh, I., Xu, T., Gui, T., Bao, Y., Li, Z.: Optical SEFDM system; bandwidth saving using non-orthogonal sub-carriers. IEEE Photonics Technol. Lett. **26**(4), 352–355 (2014)
10. Xu, T., Grammenos, R.C., Marvasti, F., Darwazeh, I.: An improved fixed sphere decoder employing soft decision for the detection of non-orthogonal signals. IEEE Commun. Lett. **17** (10), 1964–1967 (2013)
11. Kanaras, I., Chorti, A., Rodrigues, M., Darwazeh, I.: Investigation of a semidefinite programming detection for a spectrally efficient FDM system. In: 2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications, Tokyo, pp. 2827–2832 (2009)
12. Xu, T., Grammenos, R.C., Marvasti, F., Darwazeh, I.: An improved fixed sphere decoder employing soft decision for the detection of non-orthogonal signals. IEEE Commun. Lett. **17** (10), 1964–1967 (2013)
13. Zhang, H., Fang, F., Cheng, J., Long, K., Wang, W., Leung, V.C.M.: Energy-efficient resource allocation in noma heterogeneous networks. IEEE Wirel. Commun. **25**(2), 48–53 (2018)
14. Liu, X.: Spectrum Resource Optimization for NOMA-based Cognitive Radio in 5G Communications. IEEE Access
15. Benjebbour, A., Saito, Y., Kishiyama, Y., Li, A., Harada, A., Nakamura, T.: Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access. In: 2013 International Symposium on Intelligent Signal Processing and Communication Systems, Naha, pp. 770–774 (2013)
16. Li, H., Huang, Z., Xiao, Y., Zhan, S., Ji, Y.: Solution for error propagation in a NOMA based VLC network: symmetric superposition coding, vol. 25, No. 24, 27 November 2017. OPTICS EXPRESS 29856
17. Zhai, D., Zhang, R., Cai, L., Li, B., Jiang, Y.: Energy-efficient user scheduling and power allocation for noma based wireless networks with massive IoT devices. IEEE Internet Things J. (2018)
18. Saito, Y., Benjebbour, A., Kishiyama, Y., Nakamura, T.: System-level performance evaluation of downlink non-orthogonal multiple access (NOMA). In: 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), London, pp. 611–615 (2013)

Hypergraph-Coloring-Based Pilot Allocation Algorithm for Massive MIMO Systems

Yongjin Lian[1(✉)], Tiankui Zhang[1], and Youxiang Wang[2]

[1] Beijing Laboratory of Advanced Information Networks, Beijing Key Laboratory of Network System Architecture and Convergence, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
{lianyongjin,zhangtiankui}@bupt.edu.cn
[2] China Unicom, Network Technology Research Institution, Beijing, China

Abstract. Pilot contamination caused by unavoidable reuse of pilots in adjacent cell limits the performance of the massive multiple-input multiple-output (MIMO) systems severely. In this paper, we propose a hypergraph-coloring-based pilot allocation (HCPA) algorithm for pilot contamination mitigation, which can improve the system sum rate with the minimum achievable data rate requirement. In the proposed algorithm, an interference relationship graph with hyperedges to restrict the minimum uplink achievable rate is established, and then the pilots are allocated by graph coloring method. The numerical results show that the proposed algorithm can improve both the minimum achievable data rate and the total uplink sum rate.

Keywords: Massive MIMO · Pilot contamination · Pilot allocation · Hypergraph coloring

1 Introduction

Massive multiple-input multiple-output (MIMO) has been recognized as a promising technology to meet the demand of high data capacity for the next-generation mobile network [1]. By using a very large number of antennas at the base station (BS), massive MIMO can achieve several advantages theoretically, such as increasing the system capacity, improving the energy and spectral efficiencies, and so on [2]. However, the limited orthogonal pilot resource in massive MIMO systems leads to the reuse of pilot sequences among different cells, which will generate severe inter-cell interference to influence the system performance [3]. This phenomenon is called *pilot contamination*.

In the previous studies, several methods have been proposed to mitigate pilot contamination. By asynchronously transmitting among adjacent cells, the time-shifted pilot scheme can partition the cells into several groups and exploit a

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 811–819, 2019.
https://doi.org/10.1007/978-981-13-6264-4_95

time-shifted protocol within each group, which leads to the reduction of pilot contamination [4]. Based on the second-order statistics of channel state information, an angle-of-arrival (AOA)-based method is proposed to assign the same pilot to users with non-overlapping AOA mutually [5]. However, the above algorithms suffer from high complexity. In contrast, to achieve near-optimal performance with consideration of the required computational complexity, a new method based on the cross-entropy optimization (CEO) framework is proposed to decrease both the pilot contamination effect and the computational load [6]. In the meanwhile, a weighted-graph-coloring-based pilot decontamination (WGC-PD) scheme is proposed that aims to mitigate the pilot contamination by efficiently allocating pilots among users in the interference graph [7]. On the other hand, a pilot allocation algorithm in [8] utilizes the ability of acquiring and tracking the large-scale fading coefficients between the users and the BSs to maximize the minimum achievable data rate.

In this paper, we proposed a hypergraph-coloring-based pilot allocation (HCPA) algorithm. In contrast to the conventional pilot allocation algorithm, the proposed algorithm harnesses knowledge of the large-scale coefficients of the users' channel to reduce pilot contamination. First, the pilot allocation issue is formulated as a combinational optimization problem to maximize the system sum rate with the minimum achievable data rate requirement. Then, we construct a hypergraph according to independent interference and cumulative interference where we adopt a more reasonable independent interference relationship than existed graph coloring methods and introduce hyperedges to restrict the minimum achievable data rate. Finally, we propose a hypergraph coloring method to achieve the goal of pilot allocation.

2 System Model

We consider an uplink multi-cell TDD massive MIMO system composed of L hexagonal cells. Each cell has a BS with M antennas and $K\,(M \gg K)$ single-antenna users. The channel matrix $\mathbf{H}_{ij} \in \mathbb{C}^{M\times K}$ between K users in the j-th cell and M antennas of the BS in the i-th cell can be represented as

$$\mathbf{H}_{ij} \triangleq [\mathbf{h}_{ij1}, \mathbf{h}_{ij2}, \cdots \mathbf{h}_{ijK}], 1 \le i \le L, 1 \le j \le L, \tag{1}$$

where $\mathbf{h}_{ijk} = \mathbf{g}_{ijk}\sqrt{\beta_{ijk}} \in \mathbb{C}^{M\times 1}$ denotes the channel gain between the k-th user in the j-th cell and the BS in i-th cell, $\mathbf{g}_{ijk} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ denotes the corresponding small-scale fading vector, and β_{ijk} denotes the large-scale fading coefficient. We assume that the large-scale fading coefficients change slowly during several coherence intervals, in the meanwhile, they are independent over the antenna index and known to everyone.

Due to the limited coherence time, K pilot sequences $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots \mathbf{s}_K] \in \mathbb{C}^{\tau\times K}$ $\left(\mathbf{s}_i \in \mathbb{C}^{\tau\times 1}\right)$ are reused in adjacent cells, while different users in each cell use orthogonal pilots to avoid severe intra-cell interference, so we assume that

$\mathbf{S}\mathbf{S}^H = \mathbf{I}_K$. Here, τ denotes the pilot length and $(\cdot)^H$ denotes the Hermitian transpose. The received signal $\mathbf{Y}_i \in \mathbb{C}^{M\times\tau}$ at the BS in the i-th cell is

$$\mathbf{Y}_i = \sqrt{p_p}\sum_{j=1}^{L}\mathbf{H}_{ij}\mathbf{s}_j^T + \mathbf{N}_i, \tag{2}$$

where p_p denotes the pilot transmit power, $\mathbf{N}_i$ is the independent and identically distribution additive white Gaussian noise.

In this paper, we assume that minimum mean square error (MMSE) estimation method and zero-forcing (ZF) linear detector are used. According to [9], we can get the signal-to-interference ratio (SINR) of the k-th user of the i-th cell, which can be calculated as follows:

$$SINR_{ik} = \frac{\left|\mathbf{a}_{ik}^H\mathbf{h}_{iik}\right|^2}{\sum\limits_{j\neq k}\left|\mathbf{a}_{ik}^H\mathbf{h}_{iij}\right|^2 + \sum\limits_{l\neq i}\sum\limits_{m=1}^{K}\left|\mathbf{a}_{ik}^H\mathbf{h}_{ilm}\right|^2 + \left\|\mathbf{a}_{ik}\right\|_2^2/p_{ul}} \to \frac{\beta_{iik}^2}{\sum\limits_{l\neq i}\beta_{ilk}^2}\,(M\to\infty), \tag{3}$$

where $\mathbf{a}_{ik}$ represents the linear detection vector of the k-th user in the i-th cell and $\sum_{l\neq i}\beta_{ilk}$ denotes pilot contamination caused by pilot reuse. According to [4], we can calculate the uplink achievable data rate of the k-th user in the i-th cell. The formula can be expressed as

$$R_{ik} = (1-\mu)\,\mathrm{E}\left\{\log_2\left(1+SINR_{ik}\right)\right\}, \tag{4}$$

where μ evaluates the loss of spectral efficiency caused by the uplink pilot transmission, which is actually the ratio of the pilot sequences length τ and the channel coherence time T , i.e., $\mu=\frac{\tau}{T}$.

3 Hypergraph-Coloring-Based Pilot Allocation Algorithm

3.1 Problem Formulation

Most of the researches aim to maximize the total uplink achievable rate of all KL users in L hexagonal cells, while they ignore the user which contains the minimum uplink achievable data rate. Therefore, we propose a method to maximize the total uplink achievable rate with ensuring communication quality of all users in the system, which can be formulated as the following optimization problem:

$$\begin{aligned}&\max_{\mathcal{F}_s}\sum_{i=1}^{L}\sum_{k=1}^{K}\log_2\left(1+\frac{\beta_{iik}^2}{\sum_{j\neq i,p_{jk'}=p_{ik}}\beta_{ijk'}^2}\right)\\ &s.t.\quad \log_2\left(1+\frac{\beta_{iik}^2}{\sum_{j\neq i,p_{jk'}=p_{ik}}\beta_{ijk'}^2}\right)\geq\gamma_{th},\end{aligned} \tag{5}$$

where $\{\mathcal{F}_s : s = 1, 2, \cdots K!\}$ denotes all possible $K!$ kinds of pilot allocation ways and γ_{th} is introduced to restrict the users' minimum uplink achievable data rate.

3.2 Hypergraph Construction

For a system consisting of hexagonal cells, users at the junction of three neighboring cells may be severely affected by the inter-cell interference. To increase the minimum uplink achievable data rate, we will introduce hypergraph method to solve the pilot allocation problem. Hypergraph is a generalized graph, in which edges consist of any subset of the given set of vertices instead of exactly two vertices defined in the traditional graph [10]. The first step is to construct the hypergraph for the mutual interference between the users which are allocated the same pilot, and the next step is to color the constructed hypergraph. Finally, pilot sequences are assigned to each user one by one.

In the hypergraph construction, we define two kinds of interferers. One is independent interferer and the other is cumulative interferer.

1) Independent interferer recognition: If a pilot is only allocated to two users in different cells, the sum of uplink achievable rate of the k-th user in the i-th cell and the k'-th user in the j-th cell can be presented as

$$R_{sum} \propto \log_2\left(1+\frac{\beta_{iik}^2}{\beta_{ijk'}^2}\right)+\log_2\left(1+\frac{\beta_{jjk'}^2}{\beta_{jik}^2}\right)=\log_2\left(\left(1+\frac{\beta_{iik}^2}{\beta_{ijk'}^2}\right)\left(1+\frac{\beta_{jjk'}^2}{\beta_{jik}^2}\right)\right). \quad (6)$$

It is clear that the larger the R_{sum} is, the smaller the interference between the two users is. Therefore, we define $\eta_{ikjk'}$ to measure the intensity of independent interference between the k-th user in the i-th cell and the k'-th user in the j-th cell, which is represented as

$$\eta_{ikjk'} = \frac{1}{\left(1+\frac{\beta_{iik}^2}{\beta_{ijk'}^2}\right)\left(1+\frac{\beta_{jjk'}^2}{\beta_{jik}^2}\right)}. \quad (7)$$

Larger $\eta_{ikjk'}$ indicates user $\langle i,k\rangle$ and user $\langle j,k'\rangle$ should be allocated with different pilot sequences.

2) Cumulative interferer recognition: After all the independent interferers are determined, the next step is to find the cumulative interferes and construct the hyperedges. The cumulative interferers are gathered from more than one user. Considering the time complexity and interference circumstances, we construct every hyperedge which consists of three vertices. We define η' to indicate minimum uplink achievable data rate approximately among three users in different adjacent cells, i.e.,

$$\eta' = \min\left(\log_2\left(1+\frac{\beta_{iik}^2}{\beta_{ijk'}^2+\beta_{ilk''}^2}\right), \log_2\left(1+\frac{\beta_{jjk'}^2}{\beta_{jik}^2+\beta_{jlk''}^2}\right), \log_2\left(1+\frac{\beta_{llk''}^2}{\beta_{lik}^2+\beta_{ljk'}^2}\right)\right). \quad (8)$$

Smaller η' indicates that there will be a user having smaller uplink data rate if the three users in different cells are allocated with the same pilot sequence. To increase the minimum uplink achievable data rate, γ_{th} is introduced to determine whether three users in different cells can reuse the same pilot. When $\eta' < \gamma_{th}$, the corresponding three vertices form a hyperedge.

Mathematically, the hypergraph can be interpreted as an undirected weighted graph $G = (V, E, w)$. Vertices in set V denote all users in the system, i.e., $V = \{\langle i, k\rangle : 1 \le i \le L, 1 \le k \le K\}$, and hyperedges in set E which is a family of subsets e of V consist of two parts. The first part E_1 consists of hyperedges with two users that has a positive number $w(e)$ associated with each hyperedge, and $w(e)$ denotes the potential pilot contamination among users. The other part E_2 consists of hyperedges with three users where all weights are 1, and E_2 denotes the restrictions on low-rate users.

3.3 Proposed HCPA Algorithm

The hypergraph $G = (V, E, w)$ can be divided into $G_1 = (V, E_1, w)$ and $G_2 = (V, E_2, 1)$ according to independent interferers and cumulative interferers. For G_1, its edges consist of two vertices with positive weight η. Larger weight between two users indicates that severe interference will be introduced when the same pilot is allocated to them. Therefore, our goal is to maximize the total uplink rate by making users with large interference be assigned different pilots. For G_2, its hyperedges consist of three vertices. It means that the three users in a hyperedge are not assigned the same pilot as much as possible. The hypergraph coloring algorithm we propose can be explained step by step as follows (Fig. 1):

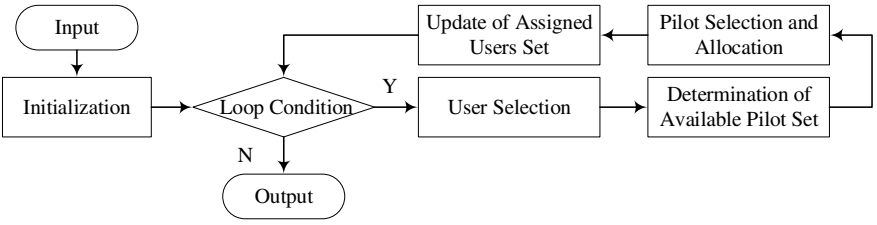


Fig. 1. HCPA algorithm flow chart

1) Input: System parameters, i.e., K, L, and the constructed hypergraph.
2) Initialization: At first, we select two users in different cells with the largest weighted edge, i.e., user $\langle i_1, k_1\rangle$ and user $\langle i_2, k_2\rangle$, then assign two pilots to them, respectively. We define Θ as set of assigned users and add the two users to Θ.
3) Loop Condition: Every user in the system will be allocated with their corresponding pilots in turn. The loop will not terminate as long as there are users who are not assigned pilots.
4) User Selection: We introduce a priority parameter $\delta_{ik} = \sum_{\langle i', k'\rangle \in \Theta, i' \ne i} \eta_{iki'k'}$ which is defined as the weight sum of the edges in E_1 connecting user $\langle i, k\rangle$ and users in other cells within Θ. Then, the user with the largest interference intensity out of the Θ will be selected, i.e., the user $\langle i_0, k_0\rangle$.

5) Determination of Available Pilot Set: To avoid the intra-cell interference, no pilot will be reused in the same cell. We define S_1 as the unassigned pilot set in the i_0-th cell. Due to the restriction of the hyperedges in set E_2, the three users in a hyperedge are not assigned the same pilot as much as possible. Therefore, we traverse all hyperedges containing the selected user $\langle i_0, k_0 \rangle$. If the other two users in the hyperedge have been assigned the same pilot, we add the pilot to set S_2. Finally, we define the set S_3 as available pilot set in which pilots belong to S_1 and do not belong to S_2.
6) Pilot Selection and Allocation: In the previous step, the available pilot set S_3 has been determined. We define $\lambda_s \left(\lambda_s = \sum_{\langle i,k \rangle \in \Theta, p_{ik}=s} \eta_{i_0 k_0 i k} \right)$ to describe the inference intensity between the users with pilot s in Θ and the selected user $\langle i_0, k_0 \rangle$ by assuming that the user is assigned with pilot s. However, considering a special case that S_3 is an empty set, we discuss two situations, respectively. If S_3 is not an empty set, the pilot s in S_3 having the smallest interference intensity λ_s is selected to be allocated to user $\langle i_0, k_0 \rangle$; if S_3 is an empty set, the pilot s in S_1 having the smallest λ_s is selected to be allocated to user $\langle i_0, k_0 \rangle$.
7) Update of Assigned Users Set: After any user is allocated with the pilot, it will be added to the set Θ.
8) Output: One-to-one correspondence between users and pilots.

4 Numerical Results

In this section, we evaluate the performance of the proposed HCPA algorithm by using Monte Carlo simulations. We consider a typical hexagonal cellular network with L cells, where each cell has K users with single antenna and a BS with M antennas. The system parameters are listed in Table 1. The large-scale fading coefficients are generated according to [7].

Table 1. Simulation parameters

Parameters	Value
Number of cells L	7
Number of users in each cell K	8
Number of BS antennas M	[32, 512]
Threshold γ_{th}	2 bps/Hz
Cell radius R	500 m
Loss of spectral efficiency μ	0.1
Path loss exponent α	3
Log-normal shadow fading σ	8 dB

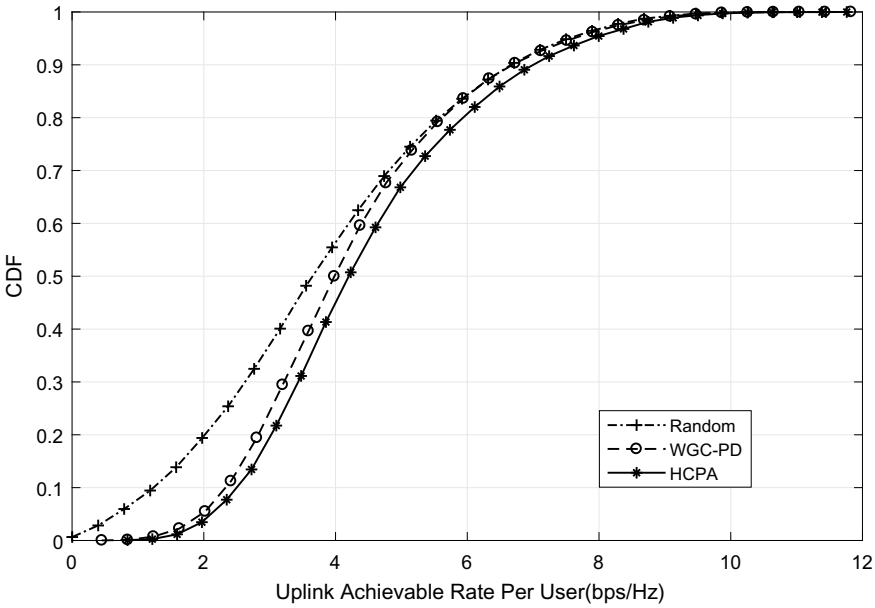


Fig. 2. CDF of uplink achievable rate per user

We compare the performance of the proposed algorithm with the performance of both: (1) the WGC-PD algorithm proposed in [7]; (2) the conventional scheme which assigns the pilots to K users randomly in each cell. Figure 2 shows the cumulative distribution function (CDF) of uplink achievable rate of all the users when M is 128. It is clearly shown that the uplink achievable rate is improved compared with the WGC-PD algorithm. This is because not only we assign orthogonal pilots to users with large interference but also adopts a more reasonable indication of the interference intensity scheme.

Figure 3 shows the average uplink achievable rate against the number of BS antennas. Due to the pilot contamination, as the number of antennas increases, the growth of average uplink rate slows down gradually. However, the growth rate proposed in this paper and in [7] is greater than the conventional allocation algorithm, because both of them mitigate pilot contamination.

Figure 4 shows the cumulative distribution function of the minimum uplink achievable data rate in each cell. It is obvious that minimum uplink achievable rate of the users in each cell becomes smaller than WGC-PD algorithm. The improvement is due to the fact that the proposed algorithm restricts the users' lower rate by the hyperedges.

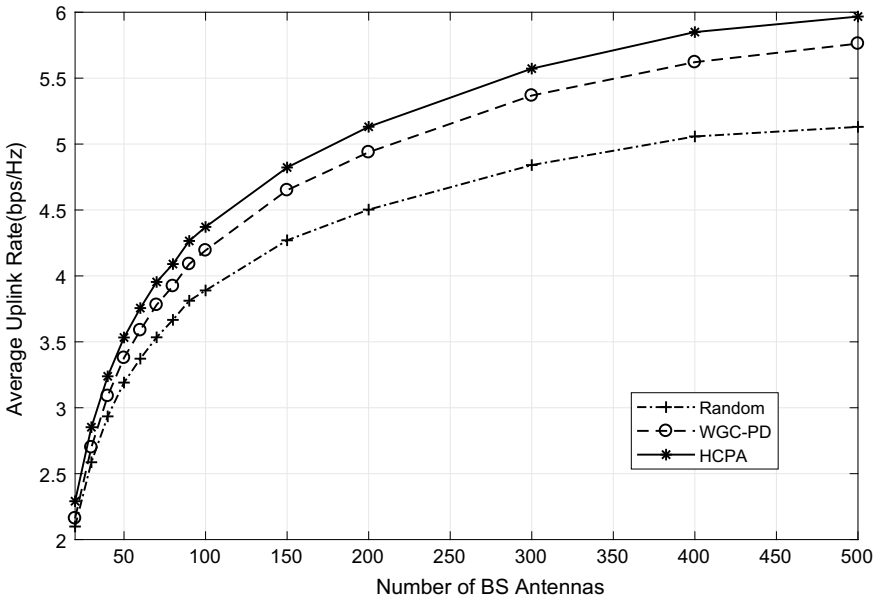


Fig. 3. Average uplink achievable rate against the number of BS antennas

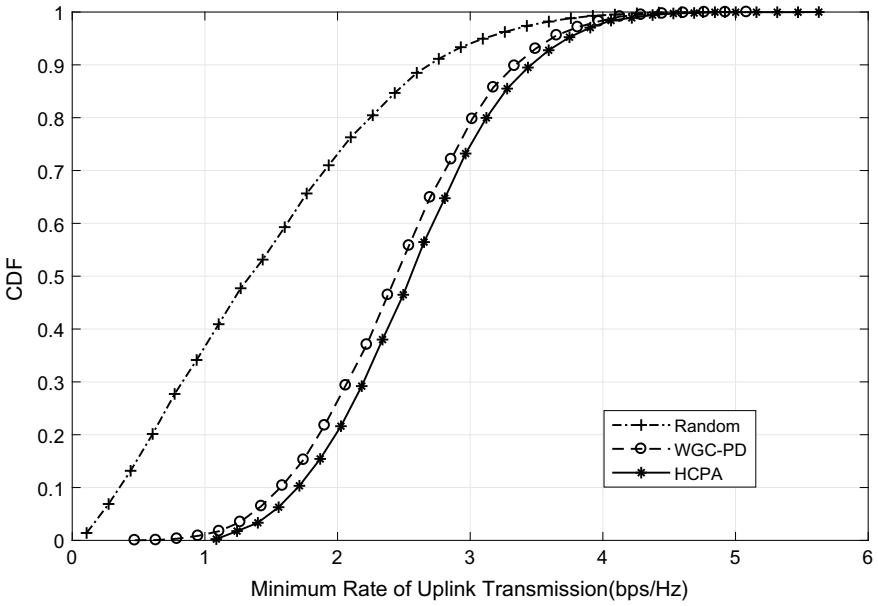


Fig. 4. CDF of minimum achievable rate per cell

5 Conclusion

In this paper, we propose a hypergraph-coloring-based pilot allocation algorithm that aims to mitigate pilot contamination in massive MIMO systems. We construct a hypergraph according to independent interference and cumulative interference where we adopt a more reasonable independent interference relationship than WGC-PD in [7] and introduce hyperedges to restrict the minimum achievable data rate. According to the constructed hypergraph, we propose the HCPA algorithm to color the hypergraph to achieve the goal of pilot allocation. Finally, our numerical results show that the proposed algorithm can improve both the minimum achievable data rate and the total uplink sum rate.

Acknowledgements. This paper is supported by National Science and Technology Major Project (No. 2016ZX03001009-003), NSF of China (No. 61502046), the 111 project (No. B17007), and Director Funds of Beijing Key Laboratory of Network System Architecture and Convergence (2017BKL-NSAC-ZJ-07).

References

1. Elijah, O., Leow, C.Y., Rahman, T.A., Nunoo, S., Iliya, S.Z.: A comprehensive survey of pilot contamination in massive MIMO-5G system. IEEE Commun. Surv. Tuts. **18**(2), 905–923 (2015)
2. Larsson, E.G., Edfors, O., Tufvesson, F., Marzetta, T.L.: Massive MIMO for next generation wireless systems. IEEE Commun. Mag. **52**(2), 186–195 (2014)
3. Lu, L., Li, G.Y., Swindlehurst, A.L., Ashikhmin, A., Zhang, R.: An overview of massive MIMO: benefits and challenges. IEEE J. Sel. Topics Signal Process. **8**(5), 742–758 (2014)
4. Fernandes, F., Ashikhmin, A., Marzetta, T.L.: Inter-cell interference in noncooperative TDD large scale antenna systems. IEEE J. Sel. Areas Commun. **31**(2), 192–201 (2013)
5. Yin, H., Gesbert, D., Filippou, M., Liu, Y.: A coordinated approach to channel estimation in large-scale multiple-antenna systems. IEEE J. Sel. Areas Commun. **31**(2), 264–273 (2013)
6. Chen, J.C., Wen, C.K., Jin, S., Wong, K.K.: A low complexity pilot scheduling algorithm for massive MIMO. IEEE Commun. Lett. **19**(10), 1842–1845 (2015)
7. Zhu, X., Dai, L., Wang, Z., Wang, X.: Weighted-graph-coloring-based pilot decontamination for multicell massive MIMO systems. IEEE Trans. Veh. Technol. **66**(3), 2829–2834 (2017)
8. Alkhaled, M., Alsusa, E., Hamdi, K.A.: A new allocation algorithm for pilot contamination mitigation in TDD massive MIMO systems. In: IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6. IEEE Press, San Francisco (2017)
9. Ngo, H.Q., Larsson, E., Marzetta, T.: Energy and spectral efficiency of very large multiuser MIMO systems. IEEE Trans. Commun. **61**(4), 1436–1449 (2013)
10. Zhang, H., Song, L., Han, Z.: Radio resource allocation for device-to-device underlay communication using hypergraph theory. IEEE Trans. Commun. **15**(7), 4852–4861 (2016)

An Efficient Hybrid Precoding Scheme for mmWave Massive MIMO Systems

Yiwei Song, Weijia Liu, and Guan Gui(✉)

Nanjing University of Posts and Telecommunications,
Nanjing 210003, People's Republic of China
{b15080234,guiguan}@njupt.edu.cn, 1042514377@qq.com

Abstract. With the explosive increase of data traffic, existing communication network fails to satisfy enormous data and high-rate communication scenarios. As alternative techniques for the next generation communication networks (5G), millimeter wave (mmWave) and massive multiple-input multiple-output (MIMO) have aroused great interest in the research community. In order to improve the hybrid precoding performance based on mmWave massive MIMO, in this paper, we propose an efficient scheme for hybrid precoding with near-optimal performance and low complexity. Here, by dividing the system data rate optimization problem into each sub-antenna optimization, it is observed that we can simply search a precoding vector close to the optimal solution. In addition, simulation results demonstrate that the proposed strategy is a near-optimal method for hybrid precoding based on the mmWave massive MIMO system, and the robustness of the proposed mean is verified among extensive simulation.

Keywords: Massive multiple-input multiple-output (MIMO) · mmWave · Hybrid precoding · Sub-antenna optimization

1 Introduction

Recently, a lot of researches have been done on 5G, and it produces some state-of-the-art technologies such as massive multiple-input multiple-output (MIMO) [1], millimeter wave (mmWave) [2], and non-orthogonal multiple access (NOMA) [3–5]. The incorporation of the mmWave and the massive MIMO techniques can achieve orders of magnitude enhancement in system throughput due to larger bandwidth [6], making it promising for future 5G wireless communication systems [7]. For massive MIMO, large-scale antennas array can enhance the system capacity and energy efficiency. Also, mmWave provides us with a candidate method to make use of the frequency band which can realize tenfold enhancement in carrier frequency.

In the last decades, many scholars have devoted to the mmWave massive MIMO system, and many efficient strategies have been derived. In [8], the author developed a mmWave massive MIMO system for enhancing system throughput.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 820–829, 2019.
https://doi.org/10.1007/978-981-13-6264-4_96

Meanwhile, the authors in [9] proposed a low-complexity hybrid analog/digital precoding for mmWave systems by employing a hybrid precoder at the transmitter and analog combiners at multiple receivers. At the same time, the energy-efficient resource allocation in two-tier massive MIMO heterogeneous networks with wireless backhaul were investigated in [10]. Furthermore, paper [11] proposed an energy-efficient weighted minimum mean square error (WMMSE)-based hybrid precoding algorithm which is proposed to maximize the achievable system sum rate. Additionally, the newly proposed geometric mean decomposition (GMD)-based scheme [12] can avoid the bit allocation issue, which can reduce the computational complexity.

However, existing methods for hybrid-precoding-based massive MIMO leads to high computational complexity, which is induced by the fact that such large-antenna system requires multiple algebraic transformation. In this paper, for the sake of promoting the performance of the mmWave massive MIMO, we propose a novel approach for hybrid precoding. Different from previous, the complicated data rate optimization issue is decomposed into many sub-antenna data rate optimization problems, which can both improve the performance in terms of hybrid precoding and decrease the computational complexity. Additionally, extensive simulations are conducted and numerical results show that the proposed hybrid precoding scheme is highly efficient and requires low computational complexity compared with traditional methods.

The rest of this paper is organized as follows. In Sect. 2, we propose a mmWave massive MIMO system. Then, to achieve high-performance hybrid precoding, a novel approach is developed based on matrix theory to achieve higher data rate, which is presented in Sect. 3. In Sect. 4, simulation results and analysis are presented for collaborating the performance of the proposed hybrid precoding, followed by the conclusions in Sect. 5.

2 System Model

First of all, as shown in Fig. 1, we consider a mmWave massive MIMO system, in which one base station (BS) with a uniform linear array (ULA) of N_t antennas and K single-antenna users are located. To be specific, it is indicated that only N RF chains can be used at the BS, and it meets the principle $N < N_t$. Also, we design a digital precoder $\mathbf{P}_D = diag[p_1, p_2, \cdots, p_N]$ in the baseband, which is a diagonal matrix for some power allocation. Here, we introduce the narrowband-ray-based channel model [13,14], and the channel matrix can be given as

$$\mathbf{H} = \zeta \sum_{\mathbf{l=1}}^{\mathbf{L}} \beta_{\mathbf{l}} \mathbf{a_r}(\theta_{\mathbf{l}}^{\mathbf{r}}) \mathbf{a_t}(\theta_{\mathbf{l}}^{\mathbf{t}}) \mathbf{g_t}(\theta_{\mathbf{l}}^{\mathbf{t}}) \mathbf{g_r}(\theta_{\mathbf{l}}^{\mathbf{r}}), \tag{1}$$

where $g_t(\theta_l^t)$ and $g_r(\theta_l^r)$ are denoted as the complex array gains of the l-th path at the BS and the user, respectively. At the same time, ζ is the normalization factor. Also, $\mathbf{a}_r(\theta_l^r)$ and $\mathbf{a}_t(\theta_l^t)$ are defined as the steering vectors at the transmit and the receive antennas. Meanwhile, $\beta_l \in \mathbb{C}$ is represented as the gains of the

l-th path. For simplicity, but without loss of generality, $\mathbf{a}_r(\theta_l^r)$ and $\mathbf{a}_t(\theta_l^t)$ can be further expressed as

$$\mathbf{a}_r(\theta_k^r) = \frac{1}{\sqrt{N_r}}[1, e^{-j2\pi\frac{d}{\lambda}\sin\theta_k^r}, \cdots, e^{-j2\pi\frac{d}{\lambda}(N_r-1)\sin\theta_k^r}]^T, \tag{2}$$

$$\mathbf{a}_t(\theta_k^t) = \frac{1}{\sqrt{N_t}}[1, e^{-j2\pi\frac{d}{\lambda}\sin\theta_k^t}, \cdots, e^{-j2\pi\frac{d}{\lambda}(N_t-1)\sin\theta_k^t}]^T. \tag{3}$$

Here, d is denoted as the antenna spacing, while λ is as the wavelength of the carrier frequency. Also, θ_k^r and θ_k^t are denoted as the azimuth (elevation) angles of departure (AOD) and angles of arrival (AoA), respectively. Then, assuming the transmitted signal vector as $\mathbf{x} \in \mathbb{C}^{N_t \times N_r}$, the received signals at the users can be presented as

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{P}_A\mathbf{P}_D\mathbf{x} + \mathbf{n}, \tag{4}$$

where $\mathbf{n}$ is the additive white Gaussian noise (AWGN) with zero mean and variance σ^2, and $\mathbf{P}_A$ is defined as the analog precoding matrix. Also, ρ is noted as the average received power. Specifically, $\mathbf{P}_A$ is consisting of N analog weighting vectors $\mathbf{a}$, and it can be formulated as

$$\mathbf{P}_A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{a}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{a}_N \end{bmatrix}. \tag{5}$$

Furthermore, for the sake of describing the hybrid precoding strategy intuitively, we denote a hybrid precoder as

$$\mathbf{P} = \mathbf{P}_A\mathbf{P}_D, \tag{6}$$

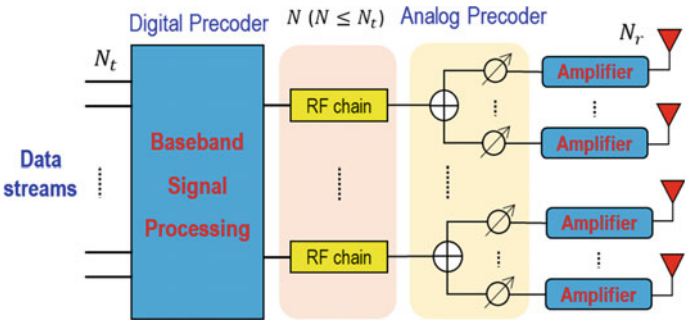


Fig. 1. The proposed mmWave massive MIMO system.

3 Proposed Hybrid-Precoding-Scheme-Based mmWave Massive MIMO

In this section, we propose a novel hybrid precoding scheme for achieving near-optimal performance based on the mmWave massive MIMO. This approach is both of low complexity and high efficiency. In addition, the robustness and computational complexity of the proposed scheme are provided in the next section.

3.1 Hybrid Precoding Architecture

In this article, based on the mmWave massive MIMO, we try to maximize the data rate of the proposed system. Here, assuming R as the data rate, we obtain

$$R = log(|\mathbf{I}_K + \frac{\rho}{N\sigma^2}\mathbf{H}\mathbf{P}\mathbf{P}^H\mathbf{H}^H|). \tag{7}$$

Here, $\mathbf{I}_K$ is a $K \times K$ identity matrix. There are two constraints existing in Eq. (8), which can be illustrated as $1/\sqrt{N_t}$ and $\|\mathbf{P}\|_F \leq N$. To be concrete, $\mathbf{P}$ should have the same amplitude and the system should meet the transmit power constraint. However, these constraints are non-convex, which brings great challenge to solve this problem. Fortunately, according to the structure of the hybrid precoding matrix, it is observed that the precoding on different sub-antenna arrays are independent. Thus, we can divide this problem into a series of sub-optimization problem.

Inspired by our consideration, Eq. (8), this problem can be rewritten as

$$\begin{aligned} R &= log(|\mathbf{I}_K + \frac{\rho}{N\sigma^2}\mathbf{H}\mathbf{P}\mathbf{P}^H\mathbf{H}^H|) \\ &= log(|\mathbf{I}_K + \frac{\rho}{N\sigma^2}\mathbf{H})[\mathbf{P}_{N-1}\mathbf{p}_N][\mathbf{P}_{N-1}\mathbf{p}_N]^H\mathbf{P}^H\mathbf{H}^H|). \end{aligned} \tag{8}$$

Using geometric transformation, we obtain

$$\begin{aligned} R &\triangleq log(|\mathbf{Q}_{N-1}|) + log(|\mathbf{I}_K + \frac{\rho}{N\sigma^2}\mathbf{Q}_{N-1}^{-1}\mathbf{H}\mathbf{p}_N\mathbf{p}_N^H\mathbf{H}^H|) \\ &\triangleq log(|\mathbf{Q}_{N-1}|) + log(1 + \frac{\rho}{N\sigma^2}\mathbf{p}_N^H\mathbf{H}^H\mathbf{Q}_{N-1}^{-1}\mathbf{H}\mathbf{p}_N). \end{aligned} \tag{9}$$

In Eq. (9), the auxiliary matrix is defined as $\mathbf{Q}_{N-1} = \mathbf{I}_K + \frac{\rho}{N\sigma^2}\mathbf{H}\mathbf{P}_{N-1}\mathbf{P}_{N-1}^H\mathbf{H}^H$. Also, it can be seen that the term $log(1 + \frac{\rho}{N\sigma^2}\mathbf{p}_N^H\mathbf{H}^H\mathbf{Q}_{N-1}^{-1}\mathbf{H}\mathbf{p}_N)$ represents the data rate of the N-th sub-antenna array. Hence, motivated by this observation, $log(|\mathbf{Q}_{N-1}|)$ is decomposed as

$$\begin{aligned} log(|\mathbf{Q}_{N-1}|) &= log(|\mathbf{Q}_{N-2}|) \\ &\quad + log(1 + \frac{\rho}{N\sigma^2}\mathbf{p}_{N-1}^H\mathbf{H}^H\mathbf{Q}_{N-2}^{-1}\mathbf{H}\mathbf{p}_{N-1}). \end{aligned} \tag{10}$$

Similarly, after N such decompositions, the data rate R can be expressed by

$$R = \sum_{i=1}^{N} log(1 + \frac{\rho}{N\sigma^2} \mathbf{p}_i^H \mathbf{H}^H \mathbf{Q}_{i-1}^{-1} \mathbf{H} \mathbf{p}_i), \tag{11}$$

where $\mathbf{Q}_i = \mathbf{I}_K + \frac{\rho}{N\sigma^2} \mathbf{H} \mathbf{P}_i \mathbf{P}_i^H \mathbf{H}^H$. Specially, $\mathbf{Q}_0 = \mathbf{I}_N$. Therefore, it is indicated that the total data rate be equal to the summation of the sub-rate of each sub-antenna array since varied precoding orders of sub-antenna arrays causes the same performance. Thus, the remaining problem is how to optimize each sub-antenna array data rate, which will be considered in the next section.

3.2 Scheme for Sub-antenna Data Rate Optimization

This subsection investigates the method for optimizing sub-antenna data rate for hybrid precoding based on mmWave massive MIMO. With the aids of the analysis above, the optimization issue of the i-th antenna array by employing i-th precoding vector $\mathbf{p}_i$ is given as

$$\begin{aligned} \mathbf{p}_i &= arg \max_{\mathbf{p}_i \in \Omega} log(1 + \frac{\rho}{N\sigma^2} \mathbf{p}_i^H \mathbf{H}^H \mathbf{Q}_{i-1}^{-1} \mathbf{H} \mathbf{p}_i) \\ &= arg \max_{\mathbf{p}_i \in \Omega} log(1 + \frac{\rho}{N\sigma^2} \mathbf{p}_i^H \mathbf{T}_{i-1} \mathbf{p}_i). \end{aligned} \tag{12}$$

Here, it is addressed that only N_t non-zero elements exist in the i-th precoding vector $\mathbf{p}_i$ and Ω is the set of all the vectors meeting the constraints stated above. Hence, Eq. (12) can be rewritten as

$$\widehat{\mathbf{p}}_i = arg \max_{\widehat{\mathbf{p}}_i \in \widehat{\Omega}} log(1 + \frac{\rho}{N\sigma^2} \widehat{\mathbf{p}}_i^H \widehat{\mathbf{T}}_{i-1} \widehat{\mathbf{p}}_i), \tag{13}$$

where $\widehat{\mathbf{T}}_{i-1}$ is the corresponding sub-matrix of $\mathbf{T}_{i-1}$ with size $N_t \times N_t$. Then, we adopt the singular value decomposition (SVD), and $\widehat{\mathbf{T}}_{i-1}$ is reformulated as

$$\widehat{\mathbf{T}}_{i-1} = \mathbf{V} \Sigma \mathbf{V}^H. \tag{14}$$

Here, $\mathbf{V}$ is an $N_t \times N_t$ unitary matrix, while Σ is a diagonal matrix that comprises the singular values of $\widehat{\mathbf{T}}_{i-1}$. Therefore, assuming $\mathbf{v}_1$ as the first right singular vector of $\widehat{\mathbf{T}}_{i-1}$, we formulate the optimization problem as

$$\widehat{\mathbf{p}}_i^{opt} = arg \min_{\widehat{\mathbf{p}}_i \in \widehat{\Omega}} \|\widehat{\mathbf{p}}_i - \mathbf{v}_1\|. \tag{15}$$

In order to maximize the data rate by selecting the precoder, the algorithm is presented in Algorithm 1.

Algorithm 1. Sub-antenna data rate optimization for mmWave massive MIMO system.

Input: $\widehat{\mathbf{T}}_{i-1}$, iteration number M, initial solution $opt^{(0)}$.
Output: $\widehat{\mathbf{p}}_i^{opt}$, $\mathbf{P}$.
1: Initialize solution $opt^{(0)} = 0$, and set the iteration number M.
2: **for** $1 \leq m \leq M$
3: $\mathbf{b}^{(m)} = \widehat{\mathbf{T}}_{i-1} opt^{(m-1)}$.
4: Introducing index $c^{(m)}$, $c^{(m)} = arg\max |b_j^{(m)}|$, where j is the column of $\mathbf{v}_1$.
5: **case 1** $(1 \leq m \leq 2)$**:** $r^{(m)} = c^{(m)}$.
6: **case 1** $(m > 2)$**:**
7: $r^{(m)} = \frac{c^{(m)} c^{(m-2)} - (c^{(m-1)})^2}{c^{(m)} - 2c^{(m-1)} + c^{(m-2)}}$.
8: **end if**
9: $opt = \frac{\mathbf{b}^{(m)}}{r^{(m)}}$.
10: **end for**
11: **for** $1 \leq i \leq N$
12: Calculate $\mathbf{v}_1$ and Σ of $\widehat{\mathbf{T}}_{i-1}$.
13: Using algebraic transformation, we obtain

$$\widehat{\mathbf{p}}_i^{opt} = \frac{1}{N_t} \|\mathbf{v}_1\|_1 e^{j\theta}, \tag{16}$$

14: Update $\widehat{\mathbf{T}}_i$ as

$$\widehat{\mathbf{T}}_i = \widehat{\mathbf{T}}_{i-1} - \frac{\frac{\rho}{N\sigma^2} \Sigma^2 \mathbf{v}_1 \mathbf{v}_1^H}{1 + \frac{\rho}{N\sigma^2} \Sigma}, \tag{17}$$

15: **end for**
16: **return:** $\widehat{\mathbf{p}}_i^{opt}$, $\mathbf{P}$.

4 Simulation Results and Analysis

In this section, we carry out computation simulation for evaluating the performance of the proposed hybrid precoding method. In our simulation, we consider a mmWave massive MIMO system with $N_t = 64$ antennas at the BS and $K = 32$ single-antenna users. Also, $d = \lambda/2$ and the maximum number of iterations $M = 10$. Moreover, the AOA and AOD are generated in the space with $[-\pi/3, \pi/3]$ distribution randomly. Additionally, our simulation has been conducted for 1000 times and the average value is selected as the simulation results.

Figure 2 shows the data rate against signal-to-noise ratio (SNR) of the proposed hybrid precoding scheme in the massive MIMO system, in which RF chains $N = 16$, $N = 12$, and $N = 8$ are considered. It can be seen from Fig. 2 that the data rate of the hybrid precoding is reducing with the increasing signal-to-noise ratio (SNR), and it increases sharply when the SNR is large enough. Also, we can see from the simulation results that the data rate can be enhanced when adopting more RF chains, but it is noted that the constraint $N \leq N_t$ cannot be neglected.

The performance comparison in terms of the data rate against SNR based on the proposed hybrid precoding scheme is shown in Fig. 3, where the system channel condition information (CSI) is considered as perfect CSI, imperfect CSI with $\eta = 0.9$, imperfect CSI with $\eta = 0.7$, imperfect CSI with $\eta = 0.6$, and imperfect CSI with $\eta = 0.2$, respectively. Here, η is denoted as the CSI accuracy. It can be seen from Fig. 3 that the data rate is enhancing as the SNR increases. Also, it is observed that the data rate in the case of perfect CSI outperforms

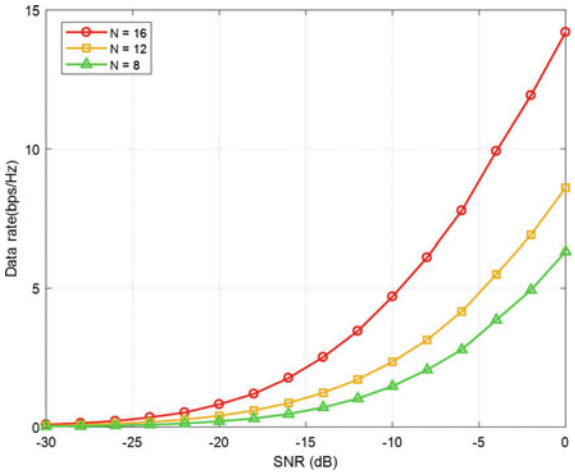


Fig. 2. Data rate of the proposed hybrid precoding method when $N = 16$, $N = 12$, and $N = 8$.

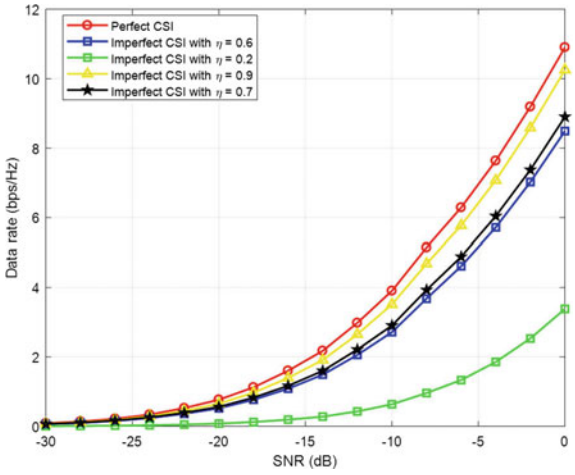


Fig. 3. Data rate with SNR in the case of perfect CSI, imperfect CSI with $\eta = 0.9$, imperfect CSI with $\eta = 0.7$, imperfect CSI with $\eta = 0.6$, and imperfect CSI with $\eta = 0.2$.

other scenarios thanks to the CSI detection that is relative to the channel state. In other words, worse channel state will worsen the system performance due to the fact that we fail to achieve high-precision CSI detection.

In Fig. 4, we depict the data rate against SNR of the proposed hybrid precoding strategy in the case of $L = 6$ with perfect CSI, $L = 6$ with imperfect CSI, and $L = 3$ with perfect CSI, $L = 3$ with imperfect CSI. Here, we set $N = 16$. As shown in Fig. 4, it is seen that the data rate achieves better performance in

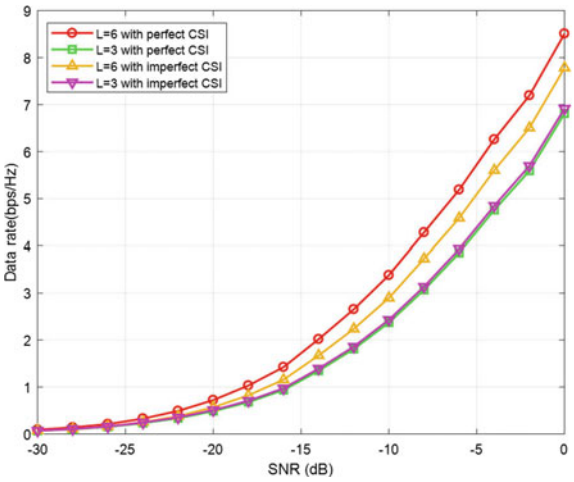


Fig. 4. Data rate with SNR in the case of $L = 6$ with perfect CSI, $L = 6$ with imperfect CSI, $L = 3$ with perfect CSI, and $L = 3$ with imperfect CSI.

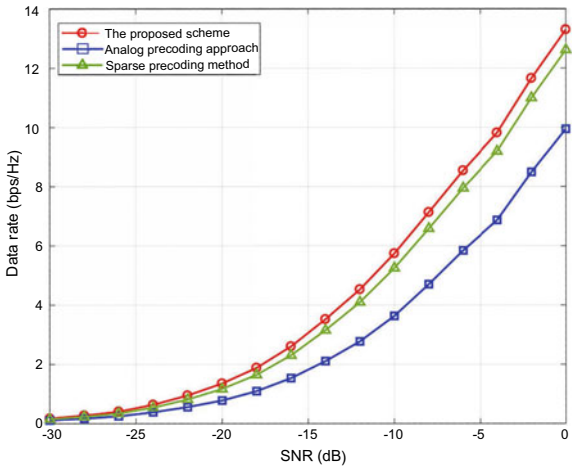


Fig. 5. Comparison of the data rate of the proposed scheme, SBEM scheme, analog precoding approach, and sparse precoding scheme.

the case of $L = 6$, and the system performance is superior in the case of perfect CSI. Thus, we can conclude that deploying more propagation paths can elevate the hybrid precoding of the proposed scheme in the massive MIMO system.

Figure 5 exhibits the data rate against the SNR of the proposed scheme, analog precoding approach [16], and sparse precoding scheme [15], respectively. We can observe that the data rate can be improved with higher SNR among all the methods. Particularly, the proposed hybrid precoding scheme has about one order of magnitude enhancement in terms of the data rate compared with the analog precoding approach, which benefits from the proposed scheme leverages of the algebraic features and the solution is more close to the optimal solution without requiring additional computational complexity. Also, the data rate performance of the proposed scheme outperforms that of the parse precoding scheme, indicating that the proposed method is highly efficient.

5 Conclusions

In this paper, we have proposed a novel hybrid precoding with sub-connected architecture for mmWave MIMO systems. To begin with, we provide a method that decomposes the total data rate optimization problem with non-convex constraints into a series of sub-antenna optimization problems. Then, it is proved that we can achieve excellent performance in terms of data rate by choosing the optimal precoder, which enhances the system capacity and reduces the computational complexity. Additionally, simulation results demonstrate that the proposed scheme can realize better performance in hybrid precoding compared with conventional methods.

References

1. Hoydis, J., ten Brink, S., Debbah, M.: Massive MIMO in the UL/DL of cellular networks: how many antennas do we need? IEEE J. Sel. Areas Commun. **31**(2), 160–171 (2013). Feb.
2. Rappaport, T.S., MacCartney, G.R., Samimi, M.K., Sun, S.: Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design. IEEE Trans. Commun. **63**(9), 3029–3056 (2015). Sep.
3. Dai, L., Wang, B., Yuan, Y., Han, S., Chih-Lin, I., Wang,Z.: Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends. IEEE Commun. Mag. **53**(9), 74–81 (2015)
4. Ding, Z., Lei, X., Karagiannidis, G., Schober, R., Yuan, J., Bhargava, V.: A survey on non-orthogonal multiple access for 5G networks: research challenges and future trends. IEEE J. Sel. Areas Commun. **35**(10), 2181–2195 (2017)
5. Huang, H., Xiong, J., Yang, J., Gui, G., Sari, H.: Rate region analysis in a full-duplex-aided cooperative nonorthogonal multiple-access system. IEEE Access **5**, 17869–17880 (2017). Aug.
6. Bai, T., Alkhateeb, A., Heath, R.: Coverage and capacity of millimeter-wave cellular networks. IEEE Commun. Mag. **52**(9), 70–77 (2014)
7. Pi, Z., Khan, F.: An introduction to millimeter-wave mobile broadband systems. IEEE Commun. Mag. **49**(6), 101–107 (2011)

8. Gao, X., Dai, L., Han, S., Chih-Lin. I., Heath, R.W., Jr.: Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays. IEEE J. Sel. Areas Commun. **34**(4), 998–1009 (2016)
9. Alkhateeb, A., Leus, G., Heath, R.W.: Limited feedback hybrid precoding for multi-user millimeter wave systems. IEEE Trans. Wirele. Commun. **14**(11), 6481–6494 (2015). Nov.
10. Yu, Q., Zhai, X., Zhao, M.: An energy-efficient hybrid precoding algorithm for multiuser mmWave massive MIMO systems. In: Proceedings of IEEE 86th Vehicular Technology Conference (VTC-Fall), Toronto, ON, pp. 1–5 (2017)
11. Hao, W., Zeng, M., Chu, Z., Yang, S., Sun, G.: Energy-efficient resource allocation for mmWave massive MIMO HetNets with wireless backhaul. IEEE Access **6**, 2457–2471 (2018)
12. Chen, C.-E., Tsai, Y.-C., Yang, C.-H.: An iterative geometric mean decomposition algorithm for MIMO communications systems. IEEE Tran. Wirel. Commun. **14**(1), 343–352 (2015). Jan.
13. Brady, J., Behdad, N., Sayeed, A.M.: Beamspace MIMO for millimeter-wave communications: system architecture, modeling, analysis, and measurements. IEEE Trans. Antennas Propag. **61**(7), 3814–3827 (2013). Jul.
14. Amadori, P.V., Masouros, C.: Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection. IEEE Trans. Commun. **63**(6), 2212–2223 (2015). Jun
15. El Ayach, O., Rajagopal, S., Abu-Surra, S., Pi, Z., Heath, R.: Spatially sparse precoding in millimeter wave MIMO systems. IEEE Trans. Wirel. Commun. **13**(3), 1499–1513 (2014). Mar.
16. El Ayach, O., Heath, R.W., Rajagopal, S., Pi, Z.: Multimode precoding in millimeter wave MIMO transmitters with multiple antenna sub-arrays. In: Proceedings of IEEE Global Communications Conference (GLOBECOM 2013), pp.3476–3480 (2013)

Effect of Satellite Attitude Errors on DBF-Based Satellite Communication

Zaiyang Jiang[1,2,3], Siyue Sun[2(✉)], Wanying Liu[1,2,3], Guang Liang[2], and Huawang Li[2]

[1] Shanghai Institute of Micro-system and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China
I.arthurj91@126.com
[2] Shanghai Engineering Center for Micro-satellites, Shanghai 201210, China
[3] University of Chinese Academy of Sciences, Beijing 100049, China
sunmissmoon@163.com

Abstract. In this paper, the influence of the satellite attitude error on pointing accuracy is discussed on the digital beamforming (DBF) technologies. The development of DBF technologies is based on analog beamforming while combining digital signal processing methods. It can use beamforming to obtain good beam directive property, and can better form beams to improve the angular resolution. Independently controllable multi-beam formation with good low minor lobe performance. Furthermore, multiple beam antenna (MBA) has been deeply studied and widely applied due to its ability to cover a large area of the ground with high gain and can also adjust the beam shape as needed.

Keywords: Digital beamforming · Satellite communication system · Multiple beam antenna · Satellite attitude error

1 Introduction

Satellite communication technology began in the 1960s, developed in the next few decades, especially when the rapid increase in demand of applications of personal mobile communications. Geostationary earth orbit (GEO) satellite system has been widely used with the development of large antenna transmitter technologies, multi-beam feed source array technologies, and DBF technologies in the last few years [1]. The GEO multi-beam satellite system is similar with the terrestrial cellular system, it covers the ground through multiple beams, furthermore, the beams are distinguished by frequency division, code division, etc. MBA has been deeply studied and widely applied due to its ability to cover a large area of the ground with high gain and can also adjust the beam shape

This work was partly supported by National Nature Science Foundation Program of China (No. 61601295), Shanghai Sailing Program (16YF1411000), and CAS Innovation Fund (No. CXJJ-16S033).

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 830–836, 2019.
https://doi.org/10.1007/978-981-13-6264-4_97

as needed [2,3]. Table 1 shows multiple beam antennas of several satellite communication systems. A multi-beam antenna is an antenna system capable of generating multiple spot beams, so as to cover a focused area on the ground. According to different demands for communication, the relationship between sub-beams and the total beam can be approximately classified as fixed areas with coverage of spot beams, and non-fixed areas with coverage of spot beams and coverage of shaped beam [4,5].

Table 1. Multiple beam antennas of satellite communication systems

Series	Giubaistar	IRIDIUM	Odyssey	MAGSS-14
Orbit height (km)	1,401	785	10,354	10,355
Antenna	Direct radiation	Direct radiation	Reflective surface	Direct radiation
Spot beam number	16	3×16	Upper 37 Lower 32	37

The multi-beam technique uses DBF techniques to calculate the synthesis coefficients based on the position of the ground receiver, the satellite position, and the satellite antenna orientation. The position of the receiver and the satellite can already be accurately obtained by positioning technologies such as global positioning system (GPS), but the current satellite attitude accuracy is limited by the cost of the satellite. A certain error of satellite attitude will further affect the performance of multi-beam communication. As a consequence, the accuracy of multi-beam communication depends on the accuracy of pointing [6,7].

2 Modeling of Spatial Geometry

The spatial reference frame should be chosen before researching the attitude of satellites. For earth oriented satellites orbiting the earth orbit, the geocentric inertial coordinate system is generally used as the reference frame of all motion, and the attitude of the satellite is described with the satellite coordinate system and the orbit coordinate system, meanwhile some auxiliary coordinate systems are also used. Definite coordinate systems as [8,9]

(1) The geocentric inertial coordinate system
Origin O_e is at the center of gravity, axis O_eZ_i is in the direction of the axis of rotation of the earth, pointing to the north pole; axis O_eX_i is pointing to the vernal point; The axis O_eY_i and the other two axes form a right-handed rectangular coordinate system, as shown in Fig. 1. This coordinate system is used to describe the state of the satellite in inertial space.
(2) The orbit coordinate system
The orbital coordinate system is the coordinate system determined by the orbital plane. The origin of coordinate O coincides with the centroid of the satellite, where the axis OZ_o points to the center of the earth, the axis OX_o

is perpendicular to the axis OZ_o in the orbital plane and points to the flight direction of the satellite. The axis OY_o vertical orbit plane and the other two axes constitute a right-handed rectangular coordinate system. As shown in Fig. 1, this coordinate system is used to describe the status of satellite orbits.

(3) The satellite coordinate system
The satellite body coordinate system is fixed on the satellite. The origin O is at the center of the satellite, the three axes are consistent with the principal axis of inertia of the satellite, and are fixed on the star body. For the Orienteering satellite, when the satellite has no attitude deviation, the system and the orbital system coincide. The axis OX_b points to the direction of satellite flight, called the roll axis. The axis OZ_b points in the geocentric direction and is called the yaw axis. The axis OY_b and the other two axes form a right-handed rectangular coordinate system, called the pitch axis. As shown in Fig. 1.

(4) North–East–Down coordinate system
Origin O can be any point on the surface of the earth, axis O_dX_θ points north along the tangent line of the geographical meridian axis. Axis O_dY_λ points east along the tangent line of geographic latitudes. Axis O_dZ_γ forms a right-handed rectangular coordinate system along the radial direction of axis O_dX_θ and axis O_dY_λ, pointing to the center of gravity, as shown in Fig. 1. This coordinate system is a coordinate system used to describe the geomagnetic field.

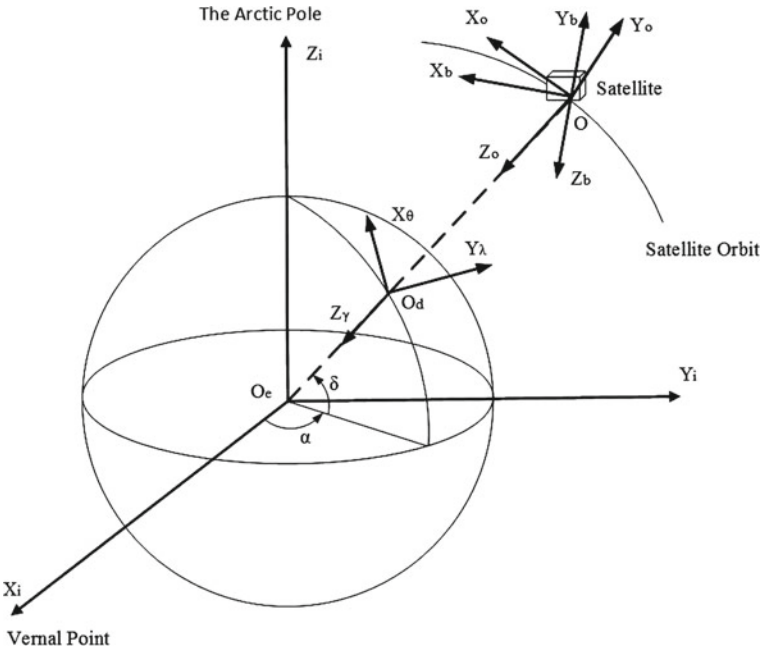


Fig. 1. Earth zone division and static device density index

(5) Other coordinate systems
In the process of satellite attitude description and control, some other coordinate systems are also involved, such as the geostationary coordinate system $O_eX_gY_gZ_g$, which is used to obtain the satellite's right longitude declination data and calculate the magnetic vector under the geographic system, and sensor coordinate system $O_mX_mY_mZ_m$ is used to describe sensor installation information, in the actual application process, the output value of the sensor needs to be converted to the satellite system through the conversion matrix.

3 Influence of the Satellite Attitude Error on Pointing Accuracy

The composition of target measurement error is relatively complex. Satellite attitude error, sensor pointing error, and image plane position error are all important parts of the measurement error. However, there is no obvious correspondence in different measurement errors. At present, the measurement error in the study is usually used as a known parameter, and the measurement error model established in the existing literature is mainly for synchronous orbit satellites. Measurement error is an important factor affecting the target positioning accuracy of space-based optical monitoring systems [10]. This paper defines the root mean

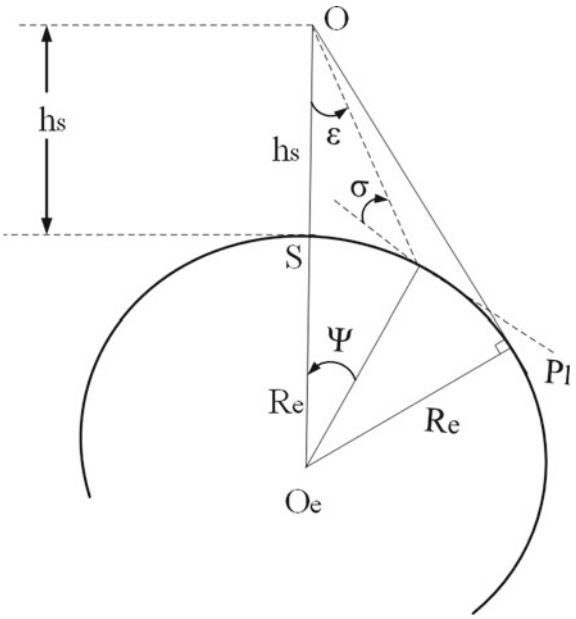


Fig. 2. Earth zone division and static device density index

square error σ of the target measurement error as

$$\sigma_{ra}^2 = [(\frac{\partial \psi}{\partial \varepsilon})^2 \sigma_\varepsilon^2 + \psi^2 \sigma_\beta^2] R_e^2, \tag{1}$$

where ψ is the vector geocentric angle of the point casted by satellite, ε is the pitch angle, σ_ε is the root mean square error of ε, β is the azimuth angle, σ_β is the root mean square error of β, and R_e is the radium of the earth, as shown in Fig. 2.

The relationship of ψ and ε can be described as

$$\frac{\sin \varepsilon}{R_e} = \frac{\sin(\varepsilon + \psi)}{R_e + h_s} \tag{2}$$

where h_e is the height of the satellite from the earth.

$$\frac{\partial \psi}{\partial \varepsilon} = \frac{(R_e + h_s)\cos \varepsilon}{\sqrt{R_e^2 - (R_e + h_s)^2 \sin^2 \varepsilon}} - 1 \tag{3}$$

4 Numerical Results and Discussions

In this section, we will see the effect of attitude error on satellite communication performance by simulation. Figure 3 compares Cramer–Rao lower bound (CRLB) of the target measurement error with different attitude errors of the pitch angle ε and the azimuth angle β. As can be seen in the figure, even the root mean square error is big enough, the CRLB is small relative to the diameter of positioning range.

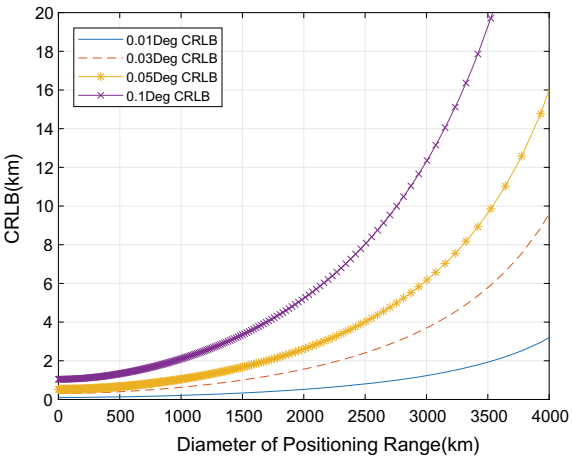


Fig. 3. Earth zone division and static device density index

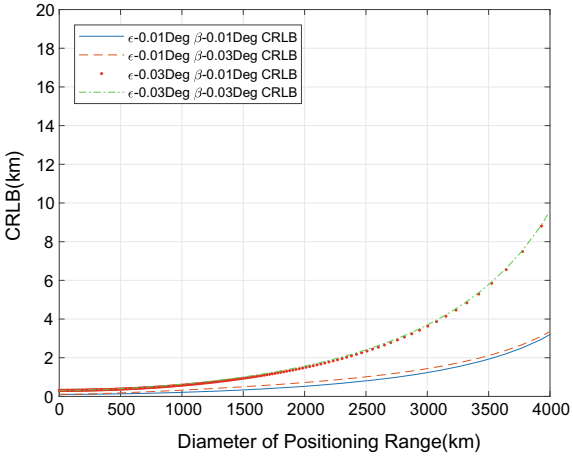


Fig. 4. Earth zone division and static device density index

Figure 4 compares Cramer–Rao lower bound (CRLB) of the target measurement error with different attitude errors of the pitch angle ε and the azimuth angle β. As can be seen in the figure, the influence of the pitch angle ε is much bigger than that of the azimuth angle β.

5 Conclusion

This paper proposed a modulation of pointing accuracy influenced by satellite attitude error, especially the pitch angle and the azimuth angle. As shown in this work, satellite attitude error is an influence which cannot be neglected on positioning.

References

1. Kechu, Y., et al.: Recent development and its prospect of satellite communications. J. Commun. **36**(6), 1 (2015)
2. Ueno, K., Itanami, T., Kumazawa, H., Ohtomo, I.: Design and characteristics of a multiband communication satellite antenna system. IEEE Trans. Aerospace Electron. Syst. **31**(2), 600–607 (1995)
3. Jilla, C.D., Miller, D.W.: Satellite design: past, present and future. Int. J. Small Satellite Eng. **8** (1997)
4. Zheng, Y., Ke, G., Sweeting, M.: Tsinghua micro/nanosatellite research and its application (1999)
5. Shamma, M., Michaelis, T.: Double gimballed momentum wheel design for small satellites. Spacecr. Guidance Navig. Control Syst. (1991)
6. Maître, H., Luo, W.: Using models to improve stereo reconstruction. IEEE Trans. Pattern Analy. Mach. Intell. **14**(2), 269–277 (1992)

7. Skullestad, A., Olsen, K., Rennehvammen, S., Fløystad, H.: Control of a gravity gradient stabilised satellite using fuzzy logic (2001)
8. Ashby, N.: The sagnac effect in the global positioning system. In: Relativity in Rotating Frames, pp. 11–28. Springer, Dordrecht (2004)
9. Schutz, B., Tapley, B., Born, G.H.: Statistical Orbit Determination. Academic Press (2004)
10. Kolitz, S.E: Passive-sensor data fusion. In: Signal and Data Processing of Small Targets 1991, vol. 1481, pp. 329–341. International Society for Optics and Photonics (1991)

Downlink Data Scheduling to Optimize the Serviceability in Fog Radio Access Networks

Zitian Zhang[1], Yue Wu[1], Yuteng Hou[2], and Jie Zhang[3(✉)]

[1] School of Information Science and Engineering,
East China University of Science and Technology, Shanghai, China
{ztzhang, yuewu}@ecust.edu.cn
[2] Beijing Railway Administration of China,
Shijiazhuang Construction Company, Shijiazhuang, China
houyl990@163.com
[3] Department of Electronic and Electrical Engineering,
University of Sheffield, Sheffield, UK
jie.zhang@sheffield.ac.uk

Abstract. In this paper, we propose to maximize the serviceability of user equipments (UEs) in fog radio access networks (F-RANs) by optimizing the downlink data scheduling for remote radio heads (RRHs) that support broadcast. To solve the optimization problem, we present a game-based heuristic algorithm. Evaluation results demonstrate the remarkable performance of our algorithm.

Keywords: Fog radio access networks · Downlink data scheduling · Serviceability of UEs

1 Introduction

The fog radio access network (F-RAN), which aims at alleviating the burden of the fronthaul links in cloud radio access network (C-RAN), is considered as a key potential technology to be implemented in the fifth generation (5G) mobile networks [1]. In F-RANs, remote radio heads (RRHs) connect to the baseband units (BBUs) in BBU pool via the fronthaul links and are deployed in close proximity to the user equipments (UEs). RRHs are equipped with local caches, thus they may store desired contents for the neighboring UEs. Controlled by a high power node (HPN), RRHs can transport data items in their local caches to the UEs directly with a high transmission rate and extremely low latency.

Although the F-RANs have remarkable advantages, they also face emerging challenges. One of the critical problems of F-RANs is the mutual interference between signals transmitted from different RRHs, as the RRHs share downlink radio resources [2]. Without a proper downlink data scheduling, such mutual interference may become so severe that it will jeopardize the system performances like throughput, spectrum efficiency, or serviceability of the UEs.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 837–842, 2019.
https://doi.org/10.1007/978-981-13-6264-4_98

There have been some initial efforts in developing downlink data scheduling mechanisms among multiple RRHs for F-RANs [3–5]. To maximize the system throughput, a superposition-coding-based downlink data delivery approach and a Hungarian-method-based downlink data scheduling algorithm were proposed in [3] and [4], respectively. A joint downlink mode selection and data scheduling algorithm were proposed to enhance the energy efficiency in [4], where both the cellular and the device-to-device (D2D) transmission modes are supported. However, none of these works has taken the broadcast ability of RRHs into consideration. As a great number of UEs in a close position are likely to require the same contents such as traffic information or live signals, RRH broadcast may obviously promote the system efficiency and be widely used in future F-RANs.

In this paper, we propose to maximize the serviceability of UEs by optimizing the downlink data scheduling for RRHs that support broadcast. We define the serviceability of UEs as the ratio between the served UEs and all the UEs that desire at least one data item from their neighboring RRHs in a scheduling period. We describe the optimization problem and present a game-based heuristic algorithm to solve it. Performance of the proposed algorithm is also evaluated through simulations.

2 Problem Description and the Proposed Algorithm

We consider an F-RAN system with N RRHs and M UEs, as shown in Fig. 1. The system consists of K downlink orthogonal frequency division multiplexing (OFDM) subchannels, each RRH has K omnidirectional transmitting antennas while each UE has one receiving antenna. We define the set of neighboring UEs of the n-th RRH as Ω_n, define the set of data items stored in the n-th RRH and the set of data items required by the m-th UE as Γ_n and Λ_m, respectively. We suppose that arbitrarily, a receiving UE in the system must be a neighbor for at least one RRH, but can be the neighbor for multiple RRHs.

We assume the HPN in Fig. 1 has a global information, it manages the N RRHs to complete the data scheduling periodically. We follow the settings in [6] that in a scheduling period, an RRH can transmit one data item to its neighboring UEs by broadcast via each downlink subchannel, and arbitrarily a UE can at most receive one data item simultaneously. Specifically, if more than one neighboring RRHs of a UE use the same downlink subchannel, the UE will be blocked at this subchannel. For each scheduling period, we consider the downlink data scheduling problem among the RRHs so that most UEs will be served.

Obviously, the downlink data scheduling problem for the F-RAN system described above is an extremely complex optimization problem. We present a game-based heuristic algorithm in this paper to solve it. In order to achieve the maximum serviceability of UEs, the choices of the data items and the subchannels, via which those data items are transmitted, for all the RRHs are interrelated. We consider the RRHs as players. The players are self-interested and each player wants to maximize the serviceability of its neighboring UEs. We define the variable i_n^k as the data item transmitted by RRH n using downlink subchannel k ($i_n^k = 0$ or $i_n^k \in \Gamma_n$), the vector

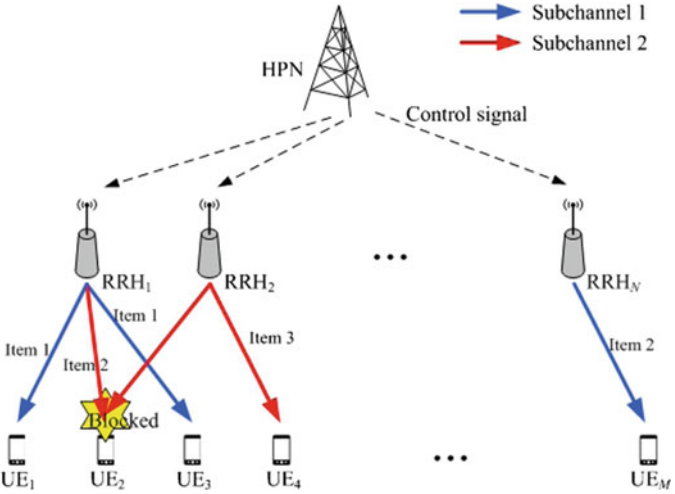


Fig. 1. System model of the F-RAN

$I_n = (i_n^1, i_n^2, \ldots, i_n^K)$ as the data items transmitted by RRH n on all the subchannels. I_n is also seen as player n's strategy. The utility function $u_n(I_n, I_{-n})$ can be defined as the number of RRH n's neighboring UEs that can be served arbitrarily by an RRH, where I_{-n} are the strategies of other players. Given I_{-n}, we first present the response method for player n to optimize its own utility function.

Algorithm 1 : Response method for player n

1: input Ω_n, Γ_n, and Λ_m for arbitrary UE $m \in \Omega_n$; input the K subchannels; input the strategies of other players, I_{-n};
2: generate the vector $I_n = (i_n^1, i_n^2, \ldots, i_n^K)$, for $\forall k = 1, 2, \ldots, K$, $i_n^k = 0$;
3: **while** ($\exists i_n^k$ in I_n equals 0 and $\Gamma_n \neq \phi$) **do**
4: given I_n and I_{-n}, calculate the set of RRH n's neighboring UEs that can not be served by any RRH, $\Omega_n^{'}$ ($\Omega_n^{'} \subseteq \Omega_n$);
5: find the data item γ belonging to Γ_n, which is desired by the most UEs in set $\Omega_n^{'}$;
6: **for** ($k = 1, 2, \ldots, K$) **do**
7: if subchannel k has not been used (i_n^k=0), suppose the data item γ is transmitted by RRH n via subchannel k and calculate the utility function of player n under this assumption, $u_{n,k}$;
8: **end for**
9: find the subchannel k^* that achieves the maximum $u_{n,k}$;
10: **if** ($u_{n,k^*} >= u_n(I_n, I_{-n})$) **then**
11: $I_n(k^*) = \gamma$, $\Gamma_n = \Gamma_n - \gamma$;
12: **else**
13: $I_n = I_n$, $\Gamma_n = \Gamma_n - \gamma$;
14: **end if**
15: **end while**

Based on the response method for a single player, the procedure of the game-based heuristic algorithm is given as follows:

Step 1: Initialize the strategy for each player. Let $I_n = (0, 0, \ldots, 0)$ for arbitrary RRH n, n = 1,…, N.

Step 2: Given other players' strategies, I_{-n}, calculate the response of player n, I_n^{new}, using the method described in **Algorithm 1**. Execute this operation for all players.

Step 3: If I_n^{new} equals I_n for arbitrarily a player, go to **Step 5**; else go to **Step 4**.
Step 4: $I_n = I_n^{new}$ for each player n. Iteration number pluses one. If the iteration number is less than a constant $\boldsymbol{Z}$, go to **Step 2**; else go to **Step 5**.
Step 5: Finish.

3 Simulation Results and Discussions

We evaluate the performance of our algorithm through Monte Carlo simulations. All results are averaged over 1000 random tests. The simulation system is shown in Fig. 2. There are six RRHs in the F-RAN and each RRH has a fixed round transmission range. In every test, the receiving UEs are uniformly distributed in the hexagon. A UE is seen as an RRH's neighbor if and only if it is in the transmission range of this RRH. Every UE in a single test randomly desires three of the total ten data items that are stored in each of the six RRHs.

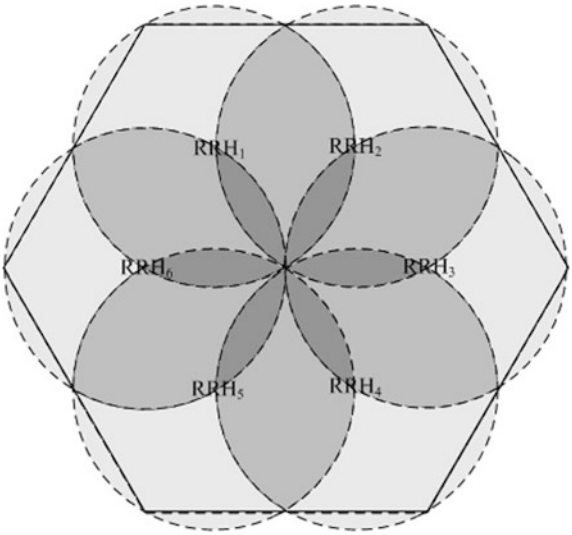


Fig. 2. The simulation system

To solve the downlink data scheduling problem, we compare our algorithm with a random allocation algorithm, in which every RRH in the F-RAN system randomly transmits K data items via the K downlink subchannels. In the simulations, the maximum iteration number of our algorithm, $\boldsymbol{Z}$, is set to be 10.

Figure 3 shows the serviceability of UEs achieved by the two algorithms under different values of M when $K = 5$. From Fig. 3, we can find that the performance of both algorithms remains stable as M increases. This is because the broadcast mechanism of RRHs lets more UEs be served corresponding to the total number of UEs. Compared with the random allocation algorithm, our algorithm can further improve the

serviceability of UEs by about 87% as it always tries to avoid the mutual interference between RRHs and makes each RRH serve the most neighboring UEs.

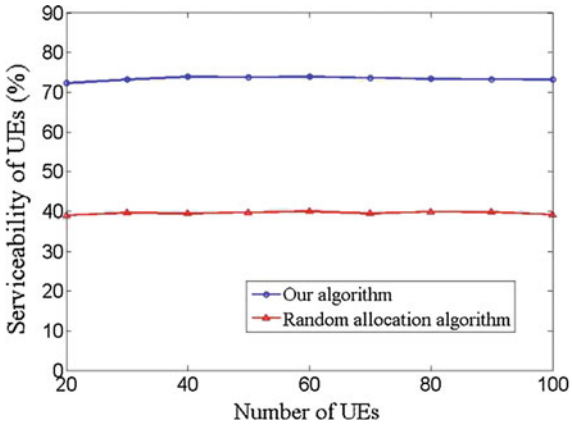


Fig. 3. Serviceability of UEs vs. the total UE number

Figure 4 plots the serviceability of UEs versus the downlink subchannel number K when there are 100 receiving UEs in the F-RAN system. The performance of the random allocation algorithm and our algorithm increases distinctly when K gets large. This can be explained as greater K may bring more chances for the RRHs to transmit data items to their neighbor UEs. When $K = 10$, our algorithm raises the serviceability of UEs by about 82% as compared with the random allocation algorithm.

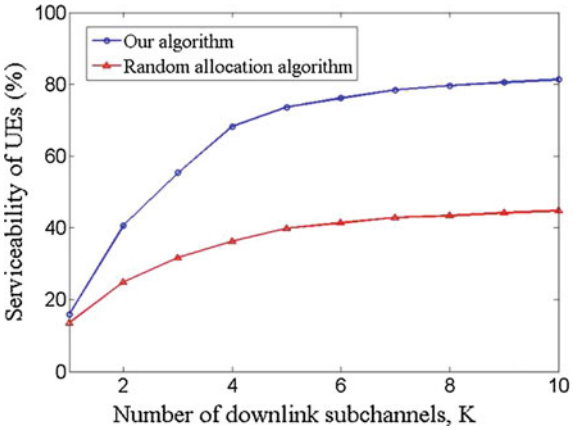


Fig. 4. Serviceability of UEs vs. the downlink subchannel number K

4 Conclusions

With the goal of optimizing the serviceability of UEs in an F-RAN system, this paper has investigated the downlink data scheduling problem for RRHs that support broadcast. We describe the system model and present a game-based heuristic algorithm. Via simulations under different scenarios, we prove that the proposed algorithm can lead to a more than 80% increase in serviceability of UEs as compared with the random allocation algorithm.

References

1. Peng, M., Yan, S., Zhang, K., Wang, C.: Fog computing based radio access networks: issues and challenges. IEEE Netw. **30**(4), 46–53 (2016)
2. Peng, M., Zhang, K.: Recent advances in fog radio access networks: performance analysis and radio resource allocation. IEEE Access **4**(99), 5003–5009 (2016)
3. Park, S.-H., Simeone, O., Shamai, S.: Joint optimization of cloud and edge processing for fog radio access networks. IEEE Trans. Wirel. Commun. **15**(11), 7621–7632 (2016)
4. Vo, D.-N., Dao, N.-N., Cho, S.: Downlink sum-rate optimization leveraging Hungarian method in fog radio access networks. In: International Conference on Information Networking (2018)
5. Xiang, H., Peng, M., Cheng, Y., Chen, H.-H.: Joint mode selection and resource allocation for downlink fog radio access networks supported D2D. In: International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (2015)
6. Liu, K., Ng, J.K.Y., Lee, V.C.S., Son, S.H., Stojmenovic, I.: Cooperative data scheduling in hybrid vehicular Ad Hoc networks: VANET as a software defined network. IEEE/ACM Trans. Netw. **24**(3), 1759–1773 (2016)

Low Earth Orbit High-Throughput Satellite Multibeam Design Schemes

Zhang Wenkai, Yang Mingchuan(✉), Yang Jiahao, and Chen Jun

Communication Research Center, Harbin Institute of Technology, 150001 Harbin, China
mcyang@hit.edu.cn

Abstract. In the low earth orbit high-throughput satellite, multibeam technology plays an indispensable role in increasing capacity and the low orbit has the remarkable advantage of low latency. A large number of researchers use the low earth orbit and high-throughput satellite constellation in the Internet of Things (IoT), in order to improve the performance of IoT service. In this paper, we focus on the multibeam design of low earth orbit high-throughput satellite. First, the common cellular structure beam design is studied and then we present a ring beam design scheme. Comparing with the cellular structure beam design, it has better performance at the center area and uses less number of spot beams to cover the same area. Based on that, we improve the ring beam design to enhance the cell edge performance and decrease the number of spot beams in coverage. We believe that the development and exploitation of these beam designs may help in making satellite constellation more attractive and competitive in the near future.

Keywords: Internet of Thing (IoT) · Low earth orbit satellite constellation · High-throughput satellite · Multibeam technology

1 Introduction

In recent years, the Internet of Things (IoT) technology has been developed rapidly and has been widely used in many fields. It is also regarded as the third wave of the development of information age following the computer and the Internet [1]. The Internet of Things needs timely access, transmission, and intelligent processing of various types of wide area geographical, environmental, spatial, and moving objects. However, current ground equipment and systems cannot provide high density or full coverage of real-time data transmission and data acquisition services for large scale or specific areas [2]. Fortunately, the use of satellite communication technology in IoT can make up for the deficiencies of the ground system. So in order to adapt to the development trend of IoT, low-latency, low-power, large-magnitude, and wide-coverage LEO satellite constellation was born. It is a large constellation of artificial satellites orbiting in LEO, almost working together to provide High-Throughput Satellite (HTS) communication to facilitate low-latency broadband IoT service to Earth [3]. Increasing companies and organizations at home and abroad are actively deploying LEO satellite constellations for IoT service, such as OneWeb constellation, SpaceX Starlink, LeoSat, Samsung, CASC Hongyan, more detailed information is shown in Table 1 [4].

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 843–851, 2019.
https://doi.org/10.1007/978-981-13-6264-4_99

Table 1. LEO satellite constellations

Constellation name	Number of satellites	Orbit (km)	Band
OneWeb	648 + 252	LEO(1200)	Ku (12–18 GHz)
			Ka (26.5–40 GHz)
Leosat	80–140	LEO(1400)	Ku (12–18 GHz)
			Ka (26.5–40 GHz)
Samsung	4600	LEO(1500)	V (40–75 GHz)
SpaceX Starlink	4425+	LEO(1110–1350)	V (40–75 GHz)
CASC Hongyan	300	LEO(–)	–

The significant increase in capacity is achieved by a high-level frequency reuse and spot beam technology which enables frequency reuse across multiple narrowly focused spot beams (usually in the order of 100s of kilometers), which both are defining technical features of high-throughput satellites [5]. Multibeam technology has been explored in almost all the satellite systems, for example, Iridium Next (48 spot beams of each satellite), Teledesic (64 spot beams of each satellite), and Skybridge (45 spot beams of each satellite) [6, 7].

The multibeam technology can generate multiple narrow spot beams at the same time so as to cover a region on the ground. Using a large number of narrow spot beams instead of global beams cannot only significantly increase the gain of satellite antennas as the beamwidth is decreasing, but also increases the capacity [8]. The larger the number of beams, the better the satellite performance, but the problems arise as the number of beams increases, the complexity of the space-borne antenna design and onboard beamforming will increase, and the satellite payload has to be larger and heavier, leading to the increasing costs and complexity of satellite design. Therefore, we need to think about some beam optimization designs to solve the problem.

In this paper, we propose the cellular structural beam design and ring beam design, and analyze their link performance. Then, based on the advantage of ring beam design, we further improve the ring beam design.

2 Multibeam Design

There are two commonly used spot beam designs: equal beamwidth design and equal area cell design. The half-power beamwidths at each beam are equal in the equal beamwidth design, and the corresponding geocentric angles are equal at each beam in the equal area cell design [9].

2.1 Cellular Structure Beam Design

According to the method of the terrestrial cellular system, the coverage area is divided into equal area cell [10], and each cell is covered by adjusting the shape of the antenna beam as shown in Figs. 1 and 2. The equal area cell spot beam design has two advantages, the half-power spot beam centered at the sub-satellite point has a larger

beamwidth and the antenna gain is lower, while the "squint" spot beam half-power beamwidth at the sub-satellite point is smaller, the higher antenna gain compensates to some extent for the increased space-free loss due to the longer propagation path. However, the antenna beamforming is difficult and it is easy to form inter-beam interference. To cover the same area, the cellular structure beam design needs large number of spot beams.

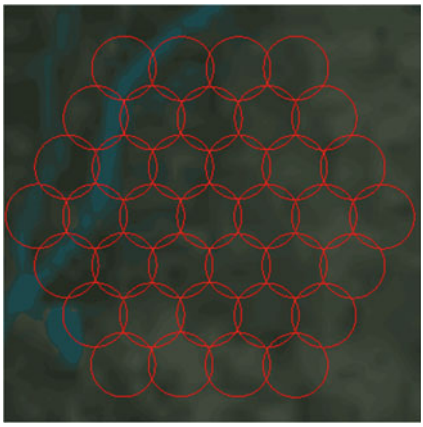


Fig. 1. Cellular beam 2D

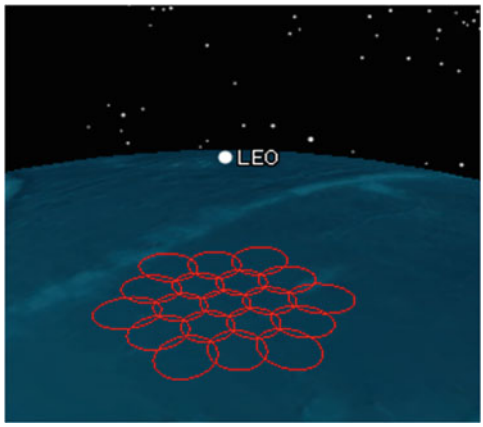


Fig. 2. Cellular beam 3D

2.2 Ring Beam Design

The fixed antenna beam size is covered with a ring, as shown in Fig. 3. This division requires a simple antenna. However, the single beam coverage of the outer ring is much larger than that of the inner ring. Excessive coverage may result in insufficient channel capacity. Then, it is improved, and the inner ring beam is properly increased and the outer ring beam is appropriately decreased, as shown in Fig. 4. The choice of beam size needs to be optimized based on parameters such as capacity, number of antennas, and beamforming difficulty.

2.3 Result Simulation and Analysis

We use two different beam designs to cover a certain area and analyze the link budget. Satellite communication links are mainly affected by free space loss, rain attenuation, and Gaussian white noise due to the line-of-sight transmissions and little effect from multipath and shadowing. Assuming only free space loss is considered and the filter is an ideal filter, the noise bandwidth is the same as the signal bandwidth. The system's parameters are set as follows (Table 2).

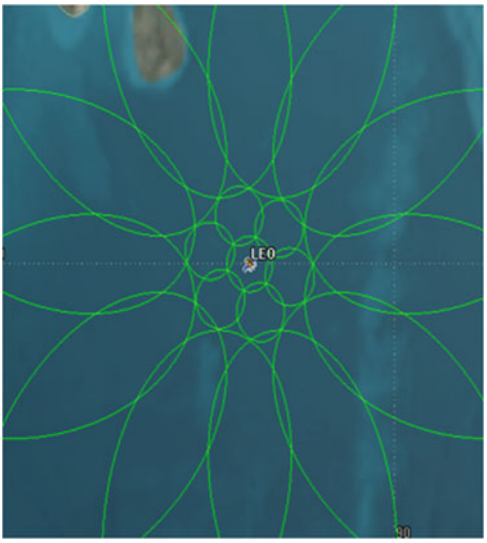


Fig. 3. Ring beam

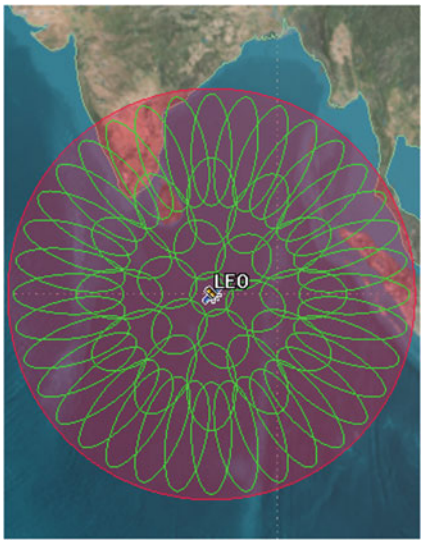


Fig. 4. Improved ring beam

Table 2. System parameters

Parameter name	Parameter value
Orbital height	700 km
Coverage radius	2000 km
Nearest distance	700 km
Longest distance	2220 km
Carrier frequency	20 GHz
Bandwidth	50 MHz
System noise	300 K
Antenna tracking loss and atmospheric attenuation	1 dB
EIRP value of terminal transmitter	50 dBw

We can calculate the free space loss by [11]

$$\mathrm{I} = \left(\frac{4\pi \cdot f \cdot d}{c}\right)^2, \tag{1}$$

where f is the carrier frequency, d is the transmission distance, and c is the speed of light. The minimum free space loss is 175.36 dB and the maximum free space loss is 185.39 dB.

The circular aperture antenna gain formulation is given by

$$G = \eta\left(\frac{\pi \cdot D}{\lambda}\right)^2 = \eta\left(\frac{\pi \cdot f \cdot D}{c}\right)^2, \tag{2}$$

where f is the carrier efficiency, D is the antenna diameter, and λ is the radiation wavelength. If the field is a parabolic distribution in the circular aperture antenna, the half-power spot beamwidth θ is

$$\theta = 72.7\frac{\lambda}{D} = 72.7\frac{C}{f \cdot D}. \tag{3}$$

We can find the relationship between antenna gain and half-power spot beamwidth θ as

$$G = \eta\left(\frac{72.7\pi}{\theta}\right)^2. \tag{4}$$

If the spot beam design is cellular, and the radius of each cell is 200 km, the maximum half beamwidth is 15.87°. The minimum half beamwidth is 5°. Assuming that the antenna efficiency is 0.55, thereby the corresponding antenna gains are, respectively, 20.94 dB and 30.60 dB. If the spot beam design is ring, and the half beamwidth of the rings distribute 14, 14, 14, 7°, 7° [12], thereby the half beamwidth of the inner ring is 14 and the half beamwidth of the outer ring is 7°. The corresponding antenna gains are, respectively, 21.65 dB and 27.68 dB.

Under clear day, the carrier-to-noise ratio of the link can be calculated by

$$\frac{C}{N} = \left(\frac{EIRP}{LI}\right)\left(\frac{G}{T}\right)\left(\frac{1}{KT}\right), \tag{5}$$

where EIRP is the carrier power, L is the sum of the antenna tracking loss and atmospheric attenuation, I is the free space loss, G is the antenna gain, T is the system equivalent noise temperature, K is the Boltzmann constant, and B is the noise bandwidth. Then, we convert the expression of the carrier-to-noise ratio into dB as given below:

$$\left(\frac{C}{N}\right) = \mathrm{EIRP(dBw)} - \mathrm{L(dB)} - \mathrm{I(dB)} + \left(\frac{G}{T}\right)\mathrm{(dB/K)} - \mathrm{K(dB/K)} - \mathrm{B(dB)}. \tag{6}$$

We can find the different performance in the assumed system as shown in Table 3 and the relationship between $\frac{C}{N}$ and EIRP for the above two different beam designs from the uplink is shown in Fig. 5.

It can be seen that there are significant differences in the carrier-to-noise ratio characteristics of the two different beam designs. The difference in the carrier-to-noise ratio between the center point and the edge of the coverage area of the cellular structure

Table 3. Two design performances

Beam design	At the center: $\frac{C}{N}$ (dB)	At the edge: $\frac{C}{N}$ (dB)	Number of spot beam
Cellular structure beam design	21.33	21.01	121
Ring beam design	22.09	18.09	55

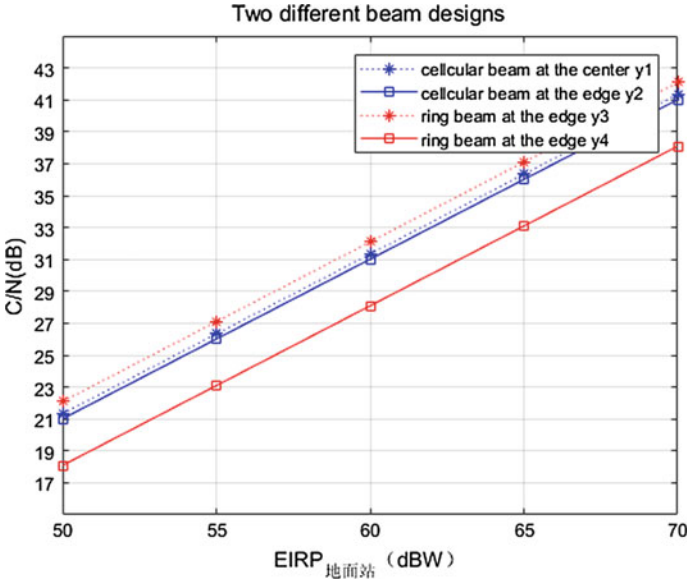


Fig. 5. Two different beam designs performance comparison

design is similar, plus the equal area cell, which is suitable for applications where the traffic distribution is even. The ring beam design has a larger difference in carrier-to-noise ratio between the center point of the coverage area and the edge, which is about 4 dB. The ring beam scheme is densely centered, but the outer ring covers a large area and is applicable to areas where the center has a large amount of traffic and the outer circle has a small amount of traffic, such as the application of the rural areas outside the central city.

The SNR of the center point of the ring beam scheme is larger than the cellular beam, and it has better coverage performance for the center service. Moreover, in order to cover the abovementioned fixed area, the number of beams of the cellular beam is more than 91, and the number of beams of the ring beam needs only 55. Therefore, we could decrease the complexity of the space-borne antenna design, the weight of payload, and the cost.

2.4 Optimization Design of Ring Structural Beam

Continue to consider achieving a better performance of the ring beam by rationally allocating beams and reducing the number of beams. Reducing the number of beams can increase the beamwidth of the antenna and reduce the gain of the antenna gain in exchange for an increase in the coverage of a single beam. Therefore, we make the 4/5 rings' coverage in the previous scheme reduced to 3 rings' coverage, and the scheme adopted is designed considering the topological structure of the beam projection. This means that the inner ring adopts one large beam cover and the half beamwidth of the above scheme is 15°; the middle ring's half beamwidth uses 18° beam coverage and the outer ring's half beamwidth uses narrow 6° to 15° beam coverage. The gap in the center ring is wider and the others are narrower. Considering the number of beams and the required antenna gain, the appropriate number of turns for the ring is 6, 7, and 8, respectively, requiring 25, 29, and 33 beams. When it is 8, the antenna gain is large and the channel quality is good. The ring beams were shown as Figs. 6 and 7, and its expression is

$$1^{\left(1^{90}_{15}\right)}2^{\left(8^{55}_{18}\right)}+3^{\left(8^{40}_{14}+16^{34}_{6.5}\right)}. \tag{7}$$

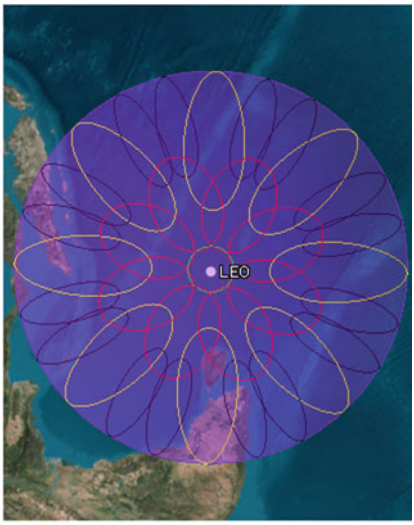


Fig. 6. 33 beams coverage 3D

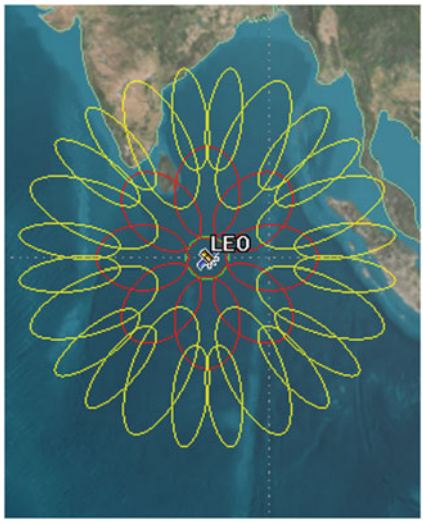


Fig. 7. 33 beams coverage 2D

The lower corner indicates the ring number, the upper mark indicates the elevation angle of the antenna, and the lower corner mark indicates the half beamwidth of the beam. For example, $1^{\left(1^{90}_{15}\right)}$ indicates that the first ring (inner ring) has one beam and the half beamwidth is 15°, the antenna elevation angle is 90°. And, $3^{\left(8^{40}_{14}\right)\left(16^{34}_{6.5}\right)}$ shows that the third ring has 24 beams, the half beamwidth and antenna elevation angle of 8 beams

are, respectively, 14°, 40° and the half beamwidth and antenna elevation angle of other 16 beams are, respectively, 6.5°, 34°.

The coverage radius of these three rings' beam design is also 2000 km. We can also find the relationship between $\frac{C}{N}$ and EIRP for in the uplink as shown below:

$$\begin{aligned}\text{At the center}: \frac{C}{N} &= \text{EIRP} - 175.36 - 1 + 21.06 - 24.8 + 228.6 - 77 \\ &= \text{EIRP} - 28.5\text{dB}, \end{aligned} \tag{8}$$

$$\begin{aligned}\text{At the edge}: \frac{C}{N} &= \text{EIRP} - 185.39 - 1 + 28.32 - 24.8 + 228.6 - 77 \\ &= \text{EIRP} - 31.27\text{dB}. \end{aligned} \tag{9}$$

Comparing to the 55 ring beams design, we can find that the SNR of 33 ring beams is similar at the center, but it has better coverage performance for the edge service as shown in Fig. 8. Moreover, the number of beams of the cellular beam is less. Therefore, we could decrease the complexity of the space-borne antenna design, the weight of payload, and the cost again.

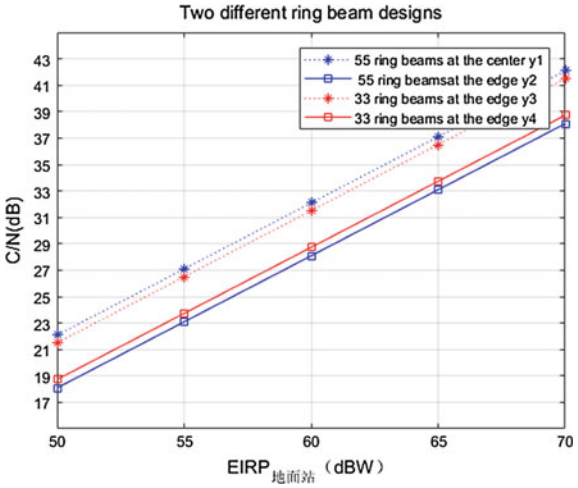


Fig. 8. Two different ring beam designs performance comparison

3 Conclusion

This paper mainly aims at improving the beam design in the Ka-band LEO satellite. First, we analyze the cellular structure beam design used by LEO satellites, and then propose a new type of ring beam design, comparing the link performance of the two designs. Next, we provide a more optimized ring beam design. This beam design not only improves the link performance in the edge area, but also significantly reduces the number of beams.

Acknowledgment. The paper is sponsored by National Natural Science Foundation of China (No. 91538104; No. 91438205).

References

1. Wang, Y.: Emerging internet of things technology-LoRa[J]. Inf. Commun. Technol. **1**, 63–68 (2017)
2. Fairhurst, G., Caviglione, L., Collini-Nocker, B.: FIRST: future internet—A role for satellite technology. In: Proceedings of IEEE International Workshop on Satellite Space Communications (IWSSC), pp. 160–164 (2008)
3. Qu, Z., Zhang, G., Cao, H., Xie, J.: LEO satellite constellation for internet of things. IEEE Access **5**, 18391–18401 (2017)
4. Dubois, T.: Eight satellite constellations promising internet service from space. Aviation Week & Space Technology. http://aviationweek.com/space/eight-satellite-constellations-promising-internet-service-space (2017)
5. Mehrotra, M.: Regulation of global broadband satellite communications. Discussion paper, ITU. pp. 29–31, (2012)
6. Iridium files for Chapter 11 bankruptcy [EB/OL]. 2002-01-02 [2015-04-16]. http://www.cnet.com/news/iridium-files-for-chapter-11-bankruptcy
7. Chan, S.P.: The birth and demise of an idea: teledesics' internet in the sky [EB/OL]. 2002-10-07 [2015-06-10]. http://community.seattletimes.nwsource.com/archive/?date=20021007&slug=teledesic070
8. Christopoulos, D., Chatzinotas, S., Zheng, G., Grotz, J., Ottersten, B.: Linear and nonlinear techniques for multibeam joint processing in satellite communications. J. Wirel. Commun. Netw. **2012**, 162 (2012). doi: https://doi.org/10.1186/1687-1499-2012-162
9. Lutz, E.: Satellite systems for personal and broadband communications [M], pp. 348–351. Springer (2000)
10. Liu, G., Wu, S.: Spot-beam design in the satellite-based cellular communication system [J]. Syst. Eng. Electron. **26**(2), 157–159 (2004)
11. Sklar, B.: Digital Communications: Fundamentals and Applications. Prentice-Hall, Englewood Cliffs, NJ (1988)
12. Almalki, F.A., Angelides, M.C.: Considering near space platforms to close the coverage gap in wireless communications: the case of the kingdom of Saudi Arabia. In: Future Technologies Conference

A Resource Allocation Scheme Based on Improved Bat Algorithm for D2D Communication System

Xujie Li[1,2(✉)], Lingjie Zhou[1], Xiaotian Deng[1], Beibei Wang[1], Chao Qiu[2], Mu Lu[3], and Chenming Li[1]

[1] College of Computer and Information, Hohai University, 210098 Nanjing, China
lixujie@hhu.edu.cn
[2] Suqian Research Institute of Hohai University, 223800 Suqian, China
[3] Jiangsu Suyuan Jierui Science & Technology Co., Ltd, 223800 Suqian, China

Abstract. In this paper, the resource allocation problem for device-to-device (D2D) communications underlaying cellular networks is formulated. In our scenario, the number of D2D user equipment is considered to be much larger than the number of mobile user equipment. Multiple D2D users can share one channel resource which is allocated to one mobile user equipment. First, a system model is established and the resource allocation problem is converted into a combinatorial optimization problem. Then, a resource allocation scheme based on the improved bat algorithm is proposed, which solves the problem of being easily getting into a local optimum solution in basic bat algorithm. Finally, the simulation results show that the performance of the proposed improved bat algorithm is better than the basic bat algorithm and the random algorithm. This scheme can provide good optimization for resource allocation in D2D communication system.

Keywords: Resource allocation · Improved bat algorithm · Capacity optimization

1 Introduction

With the increasing demand of the quality of communication services and the rapid growth of the demand for multimedia data, the lack of system spectrum resources has become an urgent problem to be solved in wireless communication at present. Device-to-device (D2D) communication is proposed as the key technology of 5G to overcome this issue [1, 2].

D2D users can communicate by reusing cellular channel resources. However, the unreasonable resource allocation strategy will introduce excessive interference. So many researchers have done a lot of work to overcome the problem. In [3], the authors mainly discuss the power control, capacity optimization, and suppression of interference in D2D communication system. In [4], a combining call admission control and power control scheme is proposed under guaranteeing QoS of every user equipment (UE). In [5], the authors consider that a cellular user channel resource is allowed to be

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 852–861, 2019.
https://doi.org/10.1007/978-981-13-6264-4_100

assigned to a single D2D user in order to void the interference of the same frequency between D2D users, but the system model is not suitable for D2D user scenario with a dense distribution. An optimal proportionally fair scheduling scheme that maximizes the logarithmic sum of the user data rates is proposed in [6]. In [7], author proposed a fully autonomous multi-agent Q-learning algorithm which has relatively fast convergence and near-optimal performance. In [8], a novel power allocation algorithm is proposed to optimize the system performance in half-duplex D2D communication underlaying full-duplex cellular networks. In [9], D2D communication system is modeled as a random network using stochastic geometry and uses a joint channel allocation and power control scheme which enables D2D links to share resources with multiple cellular users. In [10], the power emission density based interference modeling method is adopted to explore proper network settings to make sure that multiple D2D pairs can reuse the identical radio resource.

However, few papers and works had been contributed to resource allocation considering the shortage of spectrum resources. We analyze the scenario where D2D users reuse the downlink resources of cellular networks.

The main contributions of our work are as follows:

1. We model the scenario of D2D user reusing the channel resources of cellular network and obtain the objective function.
2. We propose an optimization scheme based on improved bat algorithm for D2D communications underlaying cellular networks to optimize system capacity.

The rest of this paper is organized as follows. In Sect. 2, we describe our system model. In Sect. 3, an optimization scheme based on improved bat algorithm for D2D communications is proposed and analyzed. In Sect. 4, we simulate the capacity of the system and discuss it. Finally, the conclusion is drawn out in Sect. 5.

2 System Model and Problem Formulation

2.1 System Model

The system model is assumed to work in a cellular network in the form of FDD-LTE. In general, a cellular user is assigned a single channel resource for communication. But in order to improve the spectrum efficiency of the system, the D2D communication technology is introduced into the traditional cellular mobile communication system and the D2D users can communicate by reusing the wireless communication channel resources of the cellular network. We analyze the downlink transmission in the D2D communication network as shown in Fig. 1. The scenario consists of a single base station, M cellular user equipments (CUEs), and N D2D user equipments (DUEs). DUEs are in pairs, and a pair of DUEs includes a D2D transmitting user equipment (DTUE) and a D2D receiving user equipment (DRUE). It is assumed that all user devices in the network are uniformly distributed in a circular cell with a radius of R. Here, C_i and D_j are, respectively, used to denote the ith CUE and the jth DUE.

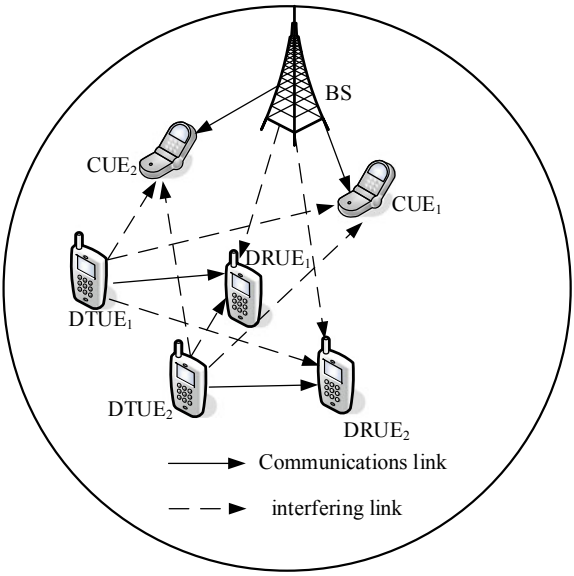


Fig. 1. System model

2.2 Problem Formulation

In the above system model, M channel resources are allocated to the N DUEs, we assume that multiple DUE pairs use channel resources of CUE_i. For example, there are four sub-channels shared by the CUEs and DUE pairs.

In D2D communication system, CUEs are interfered by signals from DUEs which reuse their channel resources. Therefore, the *SINR* of CUEs is

$$SINR_i = \frac{P_B g_{C_i}}{N_0 + \sum\limits_{D_j \in D} \delta_{ij} P_D g_{T_j,C_i}}. \tag{1}$$

Here, P_B is the transmitting power of the base station, g_{C_i} is the path gain between base station and C_i, N_0 is the noise power, P_D is the transmission power of DTUE, g_{T_j,C_i} is the path gain between $DTUE_j$ and C_i, and δ_{ij} is the reuse factor of resources between CUE_i and DUE_j.

In the D2D communication system, DRUE is interfered by signals from the base station and other DTUEs. Therefore, the *SINR* of $DRUE_j$ can be represented as

$$SINR_j = \frac{P_D g_{T_j,R_j}}{N_0 + P_B g_{R_j} + \sum\limits_{C_i \in C} \sum\limits_{D_j \in D, j' \neq j} \delta_{ij} \delta_{ij'} P_D g_{T_{j'},R_j}}. \tag{2}$$

Here, g_{T_j,R_j} is the link gain between $DTUE_j$ and $DRUE_j$, g_{R_j} is the link gain between base station and $DRUE_j$, $g_{T_{j'},R_j}$ is the link gain between $DTUE_{j'}$ and $DRUE_j$.

According to Shannon theorem, we obtain the capacity of CUEs and DUEs. The total capacity of the system is the sum of the capacity of CUEs and DUEs, that is,

$$C_{sum} = \sum_{i=1}^{M} B \cdot \log_2(1 + SINR_i) + \sum_{j=1}^{N} B \cdot \log(1 + SINR_j). \tag{3}$$

Here, B is the bandwidth of one sub-channel.

For the capacity optimization model of the system, the optimized objective function can be expressed as

$$\max\left(\sum_{i=1}^{M} B \log_2(1 + SINR_i) + \sum_{j=1}^{N} B \log(1 + SINR_j)\right) \tag{4}$$

3 Capacity Analysis of D2D Communication System Based on Improved Bat Algorithm

3.1 Description of the Basic Bat Algorithm

In 2010, bat algorithm (BA) is a new efficient bio-element heuristic algorithm proposed by Xin-Shen Yang. The main principle of the bat algorithm is to imitate bats in the preying process in order to avoid obstacles using echolocated sonar to detect prey so as to determine the changing spatial position of the prey target [11]. In the bat algorithm, individual bats in the population represent different solutions of the objective function. They choose the best solution by adjusting the loudness A, frequency f or wavelength λ, and pulse emissivity r until the condition is satisfied or target is stopped. Among them, the current position of the bat individual is judged by the size of the moderate value of the objective function to be optimized.

3.2 Improved Bat Algorithm

In view of the shortcomings of bat algorithm, such as the low precision of optimization and easy to fall into local optimum, we have improved the bat algorithm and encoded the real number of the problem into bat individuals for optimization. First, the initialization problem of the algorithm is based on the reverse learning mechanism and the population reverse initialization is adopted to generate a better initial population. Then, in order to enable bat individuals to search for better solutions in the local optimization process in the later stage of the algorithm, we refer to the cross-mutation mechanism of genetic algorithm to perform crossover and mutation operations on population bats to improve the diversity of bat individuals and avoid bat population individuals which could not achieve the expected results because they fell into partial optimum prematurely. Finally, the bat individuals are sorted by fitness values to select bat individuals with moderately favorable values as initial solutions for the next iteration. In this way,

while improving the individual diversity of the algorithm, the bat population's superior search capabilities are maintained.

There are three key steps in improved bat algorithm: the generation of the initial solutions, the generation of the new solution, and amendments of the population.

1. The generation of the initial solutions

The initial solution of the basic bat algorithm is randomly generated, which often results in the uncertainty of the results. We use the method of reverse initialization of bat populations. That is, the bat population generated by random initialization uses the Eq. (5) to generate the corresponding reverse bat population, here $x_i \in [a_i, b_i](i = 1, 2, \cdots, D)$. After combining the two solutions, calculate their fitness values, and the excellent individuals satisfying the conditions are sorted according to the fitness value as the initial bat population. The optimal bat individual is taken as the initial solution.

$$x_i' = a_i + b_i - x_i. \tag{5}$$

2. The generation of the new solution

During the local search of the algorithm, excellent bat individuals play an important role in the evolution of the population. Therefore, this paper refers to genetic algorithm and selects the excellent bat individuals in the parent bat population to perform crossover and mutation operations to generate new solutions. The crossover operation refers to selecting two individuals from the population, exchanging some of their contents in a certain way, and passing the parent's excellent genes to the offspring, so as to produce new excellent individuals. The mutation operation is to randomly select an individual from the population and select a part of the individual to mutate to produce a better individual. The purpose of doing so is to maintain the diversity of the population.

3. Amendments of the population

Since there are individuals in the new solution that do not meet the normal communication criteria, these individuals who do not meet the communication quality requirements must be excluded from the population. Then, the two bat populations before and after the crossover and mutation were mixed and ranked according to the fitness value to select a certain number of excellent individuals form a new population. Each bat represents a solution to the problem, that is, a D2D user combination method. Different solutions represent different D2D user combinations for multiplexing cellular user channel resources. Finally, the optimal bat individual is identified through the above algorithm evolution iteration. The steps for improved bat algorithm are shown in Table 1.

Table 1 Improved bat algorithm specific steps

Improved bat algorithm specific steps:
1. Parameter settings (cell radius R, number of CUEs M, number of DUEs N, iterations number $Maxge$n, etc.)
2. Reverse initialization of bat populations
3. Calculate the fitness value according to Eq. (3) to find the current optimal bat position
4. First, according to the evolution operation of basic bat algorithm to update the bat individual position, velocity, and fitness value, and then select the previous generation of excellent bat individuals according to the genetic crossover mutation to cross-mutation and update, resulting in two new solutions
5. Calculate the corresponding fitness values of the two solutions, rank them, find the current optimal solution, and select a certain number of excellent individuals which meet the conditions as the bat population for the next iteration
6. Judge whether the iteration is over. If not, return to step 4; if yes, the algorithm ends

4 Simulation Results and Performance Analysis

4.1 Simulation Establishment and Parameter Setting

The model of the system examined in this simulation is as follows: cellular users and D2D users are evenly distributed in a circular area with the base station as the center and R as the radius.

Based on this system model, due to the randomness of the user's location, it is difficult to obtain accurate values in the actual D2D communication process. Therefore, MATLAB software is used for Monte Carlo simulation. This method is suitable for computer iterative calculation. Table 2 shows simulation parameters.

Table 2 Simulation parameters

Parameter	Parameter value
Cell radius R	600 m
Cellular users M	4
D2D user logarithm N	8
SINR threshold $SINR_{th}$	4 dB
Noise power N_0	−109 dBm
Path loss index α	4
Power of BS P_B	46 dBm
Bandwidth B	0.12 MHz
Number of iterations	50

4.2 Analysis of Simulation Results

Figure 2 shows the relationship between system capacity and number of iterations. Figure 3 shows the relationship between system capacity and SINR threshold. We analyze the system capacity performance of the improved bat algorithm, basic bat algorithm, and random algorithm. It can be seen from the figure that even though the number of CUEs and DUEs in a cell is fixed, the value of the optimal system capacity for each calculation will fluctuate within a certain interval. Due to the randomness of the location of the users within the cell, it is difficult to obtain the real-time accurate value of the system capacity under the D2D communication model. It can be seen intuitively that with different values of iteration times and $SINR_{\text{th}}$, the random algorithm is the worst, the performance of the classic bat algorithm is improved, and the improved bat algorithm has the best performance. In the histogram of different parameter changes, we can obtain an optimal combination to improve the system capacity and verify that the improved bat algorithm can be applied to the D2D resource allocation field.

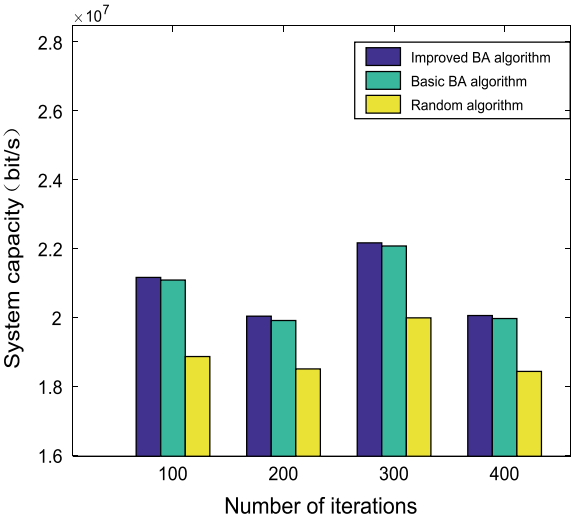


Fig. 2. Relationship between system capacity and number of iterations

Figure 4 is an iterative comparison of different algorithms. We compare the three algorithms by Monte Carlo simulation experiments. From the figure, we can see that compared with the basic bat algorithm, the bat algorithm proposed has a better initial solution to the system capacity. This is because the algorithm uses a reverse initialization earlier in the algorithm and generates better initial population. As the number of iterations increases, the speed at which the bat algorithm finds the optimal solution is much greater than that of the basic bat algorithm. Because the improved algorithm

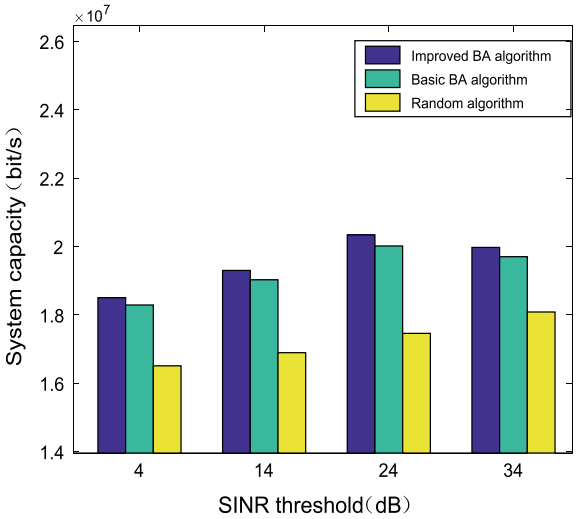


Fig. 3. Relationship between system capacity and SINR threshold

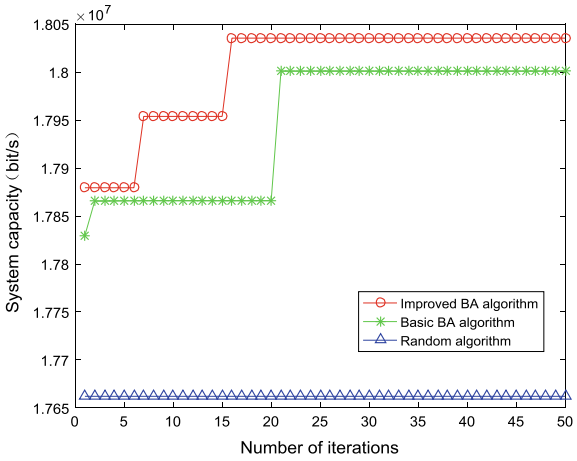


Fig. 4. Iterative comparison of different algorithms

adopts the cross-mutation operation in the middle and later stages, the population diversity is effectively improved. Compared with the basic bat algorithm, the improved bat algorithm has the good characteristics of the initial solution of the target function, the high precision of the optimization, and the fast convergence speed.

5 Conclusion

In this paper, the resource allocation problem for D2D communications underlaying cellular networks is formulated. In our scenario, we consider the number of D2D user equipment to be much larger than the number of mobile user equipment. Then, a system model is established and the resource allocation problem is converted into a combinatorial optimization problem. To improve the performance of this system, a resource allocation scheme based on the improved bat algorithm is proposed. Finally, the simulation results show that the performance of the proposed improved bat algorithm is better than the bat algorithm and the random algorithm. This result can be adopted to optimize the resource allocation in D2D communication system.

Acknowledgements. This work was supported in part by the Fundamental Research Funds for the Central Universities (No.2017B14214), Six Talent Peaks project in Jiangsu (DZXX-008), and National Natural Science Foundation of China (No. 51669014, 61601283, 61301110).

References

1. Tsiropoulos, G.I., Yadav, A., Zeng, M., Dobre, O.A.: Cooperation in 5G HetNets: advanced spectrum access and D2D assisted communications. IEEE Wirel. Commun. **24**(5), 110–117 (2017)
2. Zhang, H., Liao, Y., Song, L.: D2D-U: device-to-device communications in unlicensed bands for 5G system. IEEE Trans. Wirel. Commun. **16**(6), 3507–3519 (2017)
3. Li, X., Wang, Z., Sun, Y., Gu, Y., Hu, J.: Mathematical characteristics of uplink and downlink interference regions in D2D communications underlaying cellular networks. Wirel. Pers. Commun. **93**(4), 917–932 (2017)
4. Li, X., Zhang, W., Zhang, H., Li, W.: A combining call admission control and power control scheme for D2D communications underlaying cellular networks. China Commun. **13**(10), 137–145 (2016)
5. Zhang, R., Song, L., Han, Z., Cheng, X., Jiao, B.: Distributed resource allocation for device-to-device communications underlaying cellular networks. In: IEEE International Conference on Communications (ICC), Budapest, pp. 1889–1893 (2013)
6. Gu, J., Bae, S.J., Hasan, S.F., Chung, M.Y.: Heuristic algorithm for proportional fair scheduling in D2D-cellular systems. IEEE Trans. Wirel. Commun. **15**(1), 769–780 (2016)
7. Asheralieva, A., Miyanaga, Y.: An autonomous learning-based algorithm for joint channel and power level selection by D2D pairs in heterogeneous cellular networks. IEEE Trans. Commun. **64**(9), 3996–4012 (2016)
8. Hou, X., Liu, F., Liu, Y.: A novel power allocation algorithm for D2D communications underlaying full-duplex cellular networks. In: IEEE International Conference on Computer and Communications (ICCC), Chengdu, pp. 583–588 (2017)
9. Abdallah, A., Mansour, M.M., Chehab, A.: Joint channel allocation and power control for D2D communications using stochastic geometry. In: IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, pp. 1–6 (2018)
10. Ni, M., Pan, J.: Throughput analysis for downlink resource reusing D2D communications in cellular networks. In: IEEE Global Communications Conference (GLOBECOM), Singapore, pp. 1–7 (2017)

11. Yang, M., Wang, A., Sun, G., Liang, S., Zhang, J., Wang, F.: Signal distribution optimization for cabin visible light communications by using weighted search bat algorithm. In: IEEE International Conference on Computer and Communications (ICCC), Chengdu, pp. 1025–1030 (2017)

Design of Intelligent Exhaust System Based on LORA Communication

Yanting Xu and Yongjie Yang(✉)

School of Electronics and Information, Nantong University, Nantong 226019, China
yang.yj@ntu.edu.cn

Abstract. This article will design an intelligent exhaust system based on the characteristics of stable, practical, and efficient, which gather factory environment data by SHT20 sensors, and then use the fixed-point transmission mode of LORA to send the packaged data to relay what processes this data and transmits it to the upper computer [1]. After that, the upper computer sends the appointed threshold values of temperature and time to the relay [2] which controls the switch of the exhaust fan by judging from this threshold values. The maximum error range of SHT20 used in this design is plus or minus 4.5%, which fully meets the requirements of the system. The system has been applied to a factory in Nantong city, and the practical application shows that the system is stable, accurate, and easy to operate, which is of great significance to the realization of intelligence in the factory.

Keywords: LORA intelligent ventilation · Data collection

1 Introduction

Environmental monitoring is an important part of intelligent monitoring and control system in a factory, but in existing environmental monitoring platform, there is a general problem that temperature and humidity monitoring is not accurate and the communication is not stable. Therefore, this paper puts forward the intelligent exhaust system [3] based on LORA communication, its purpose is to design a monitoring system that can carry on the accurate monitoring of factory environment and stably transfer the data to relay, so as to control the exhaust fan.

2 Overall Design of System

In this paper, an intelligent exhaust fan control system based on LORA communication is developed, which can be divided into three parts: data collection, relay control, and upper computer display.

In Fig. 1, four data acquisition nodes which can circularly collect the change of temperature and humidity are placed in Workshop 1, Workshop 2, machine room, and distribution room of the factory separately. These acquisition nodes also have to transmit the collected data to the relay in real time every 5 s. The relay part is installed

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 862–869, 2019.
https://doi.org/10.1007/978-981-13-6264-4_101

on the switch of the exhaust fan, which has three functions: (a) Relay receives the data transmitted from the four nodes, which, in turn, takes the mean value and transmits the mean value to the upper computer for display; (b) The temperature threshold of each workshop transmitted by the upper computer is received and compared with the actual temperature in (a), respectively. The comparison results are used to control the switch of the exhaust fan to achieve the desired effect of cooling and dehumidification; and (c) The time threshold transmitted by the upper computer is received to control the working time of the exhaust fan so as to achieve the function of discharging harmful gases. The upper computer display part is mainly used to display the temperature and humidity transmitted by the relay and send the temperature threshold and time threshold to the relay.

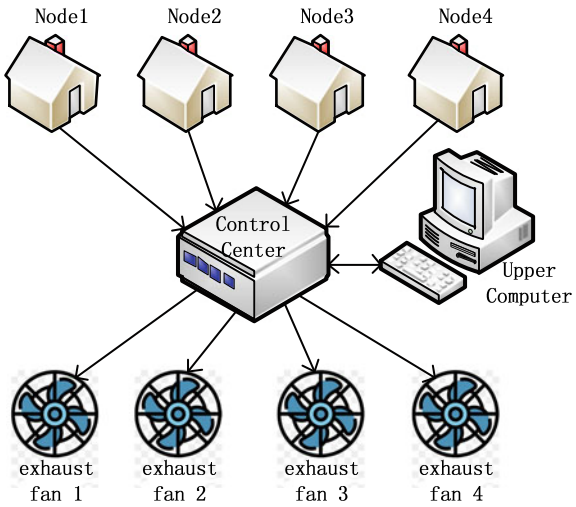


Fig. 1. System model

3 Hardware Design

The design of the hardware has two main parts: data acquisition nodes and the relay terminal. Data acquisition nodes consist of microprocessor, power supply circuit, the temperature and humidity [4] acquisition circuit, and data transmission circuit. Relay terminal consists of microprocessor, power supply circuit and relay control circuit, and data transmission circuit.

0.1 Data Acquisition Circuit

This design chooses STM32F103C8T6 [5] as the microprocessor. This microprocessor uses the high performance of ARM @ CortexTM-M3-32-bit RISC core, in which working frequency is 72 MHz built-in flash memory of as high as 128 KB and 20 KB SRAM.

As shown in Fig. 2, it is the circuit of temperature and humidity acquisition. The power supply of SHT20 is 3.3 v. A 100 nF decoupling capacitor is connected between the power supply (VDD) and the ground (GND) for filtering purpose. SCL pin is applied for communication synchronization between SHT20 and microprocessor. SDA pin is applied for the input and output of SHT20 data. SDA is valid on the SCL's ascending edge when sending commands to the sensor, and must be stable when the SCL is in high voltage. The value of SDA can be changed after the falling edge of the SCL. To avoid signal conflicts, the microprocessor can only drive SDA and SCL at low level, so an external pull-up resistor is required to pull the signal to high level, namely, R5 and R6 in Fig. 2.

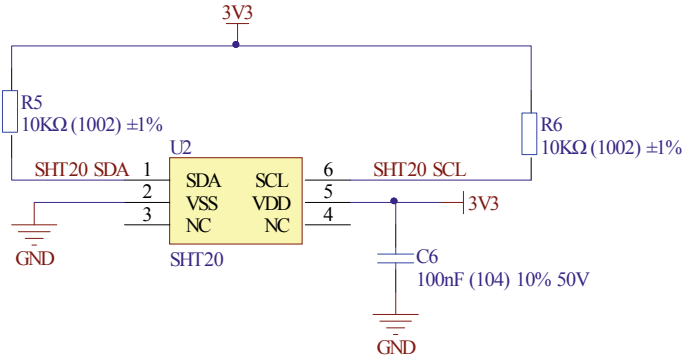


Fig. 2. The circuit of temperature and humidity acquisition

0.2 The Circuit of Data Transmission

The core of data transmission circuit is AS32-TTL-1 W, which is a 433 MHz, 1 W, high stability, industrial wireless digital transmission [6] module. It uses SX1278 as the main chip. It has the characteristics of LORA spread spectrum transmission and TTL level output. Its working frequency range is 410 MHz ~ 441 MHz. There are 32 channels with an interval of 1 M between each channel, which fully meets the communication requirements of this design.

The connection between the module and the microprocessor is shown in Fig. 3, RXD and TXD are, respectively, connected with TXD and RXD of MCU, where RXD is the input pin and TXD is the output pin. AUX is used to indicate the working status of the module. The user wakes up the external MCU and outputs the level during the online self-check initialization. MD0 and MD1 which cannot hover are connected to the IO port of MCU and determine the four working modes of the module. Mode 1 fixed-point transmission is adopted in this design, so MD0 = 1 and MD1 = 0. Fixed-point transmission is to send the data to the specified location. Its communication format is specified point address + specified point channel + the data to be sent. The specified point only receives data not addresses and channels.

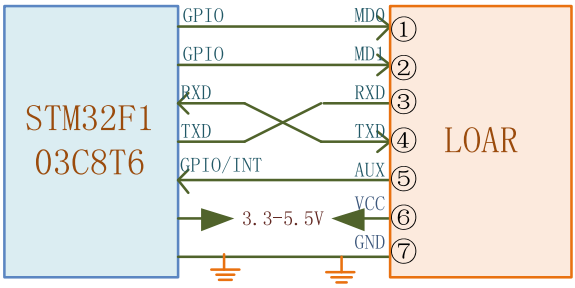


Fig. 3. The LORA module connects with the microprocessor

0.3 The Circuit of Relay Terminal

The relay chosen by relay terminal circuit is SRD-05VDC-SL-C, which is used in automatic control circuit. It is actually an automatic switch that uses a small electricity to control relative large electricity. Figure 4 is the circuit of the relay, where D3 is a light-emitting diode whose function is to indicate the relay on or off; U3, an optical coupler, isolates the front and back stages to prevent the interference level of the high voltage level of after circuit or the afferent weak current in the circuit to damage the stability of the device in the weak current. It has the effect of protection circuit; Q1 is an NPN triode, which acts as a switch in this circuit, that is, to say that controlling the coil in the relay energize or not by controlling the base electrode of the triode; D2 is a stable voltage diode, which is connected in parallel at a high voltage to achieve the effect of stable voltage.

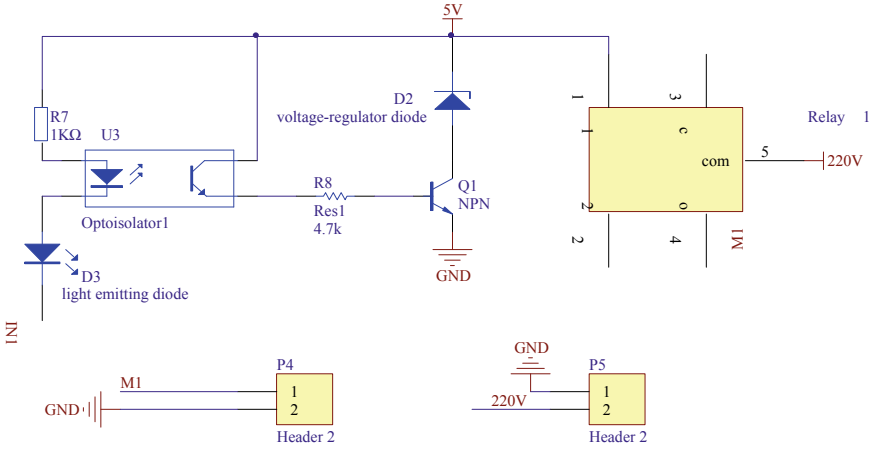


Fig. 4. Relay terminal circuit

4 Software Design

The software design part of this system is divided into two parts: the lower computer and the upper computer software.

0.1 The Software Design of Lower Computer

The lower computer software is written by Keil software, and its running process is shown in Fig. 5. The whole software includes main function, sensor driver, LORA and MCU serial port communication protocol, and subroutine of fan control. The sensor driver and the serial port communication protocol between LORA and MCU are described below.

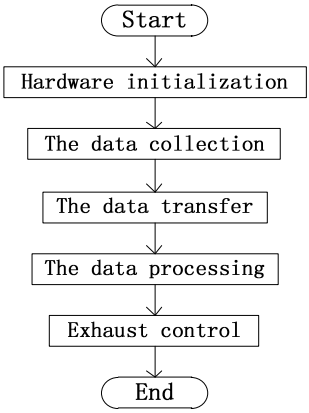


Fig. 5. Flowchart of lower computer software

0.1.1 Sensor Driver Program

The core of the sensor driver program is the IIC protocol communication which will power the sensor. It will take up to 15 ms time for the sensor to reach idle state after power up. After the transmission is started, the first IIC byte of subsequent transmission consists of a 7-bit IIC device address and an SDA directional bit. It indicates that the sensor data reception is normal by pulling down the SDA pin (ACK bit) after the falling edge of the eighth SCL clock. After sending the measurement command, the MCU must wait for the measurement to complete. The working time sequence of SHT20 operates the I/O port to complete temperature and humidity measurements.

0.1.2 LORA and MCU Serial Port Communication Protocol

After the initialization of the device is completed, hardware initialization is first carried out. Temperature and humidity data are collected at the node terminal, and the data collected are transmitted to relay separately. The upper computer will also send the lower threshold to the relay. However, the relay cannot identify the data transmitted from which node in the communication. As a result, the node adds an identification

number before the data packet when transmitting data. The data transmission format is the specified point address + the specified point channel + the node identification number + the data to be sent. The relay determines from which node the data is coming by extracting the identifier before the packet is sent.

0.2 The Software Design of Upper Computer

The upper computer software is written by Visual Studio software. As shown in Fig. 6, it is the main control interface. It includes port setting, parameter display, and threshold setting.

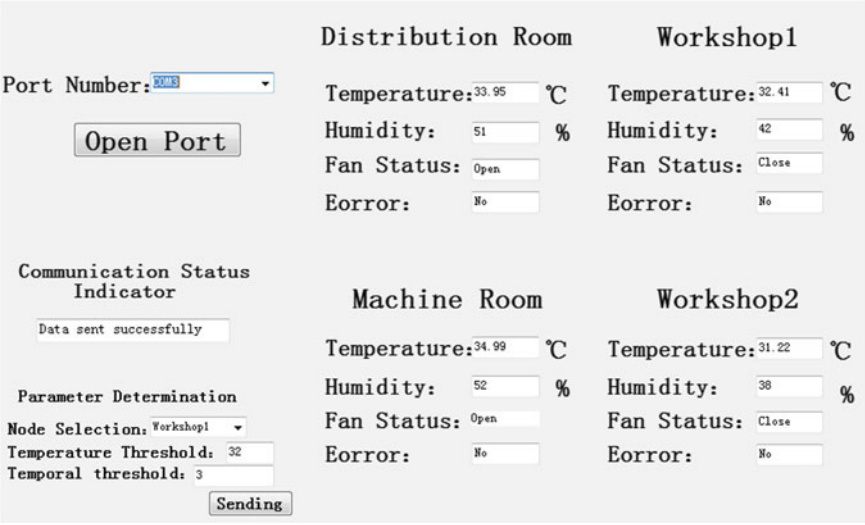


Fig. 6. Upper computer interfaces

As shown in Fig. 7, the upper computer software is mainly used to monitor the temperature and humidity of the terminal, display the data that comes from relay in real time and save it to the database to query the historical data. At the same time, the upper computer will also realize the function of thresholds setting. The user sets the appropriate threshold at the upper computer and clicks the send button to send it to the relay. At this point, if the lower computer receives the data, it will send a response signal. If the upper computer receives the response signal, "Data sent successfully" will be displayed. Otherwise, upper computer will send the threshold again. "The signal failed to send, please resend the data" will be displayed if the upper computer repeats three times without receiving the response signal.

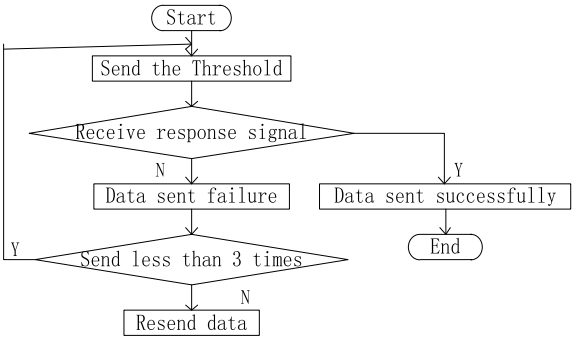


Fig. 7. The flowchart of upper computer software

5 System Test

The principle, hardware design, and software design of the whole system have been introduced above, and the data collection of the system will be introduced below to prove the feasibility of the system.

Take Workshop 1 as an example, it has to start up the system first, power on the node of Workshop 1 and relay and open the upper computer software. Afterward, it is necessary to set the temperature threshold of 32 and click the send button to send it to relay. For the convenience of illustration, data with different two states of the exhaust fans are selected for illustration. The parameters and states are shown in Table 1:

Table 1. System data

Time	Temperature (°C)	Humidity (%)	Threshold	State
10:30	30.36	43	32	关
10:35	31.51	41		关
10:40	32.01	48		开
10:45	32.98	50		开

As can be seen from the table, when the mean temperature of each node is greater than the threshold value, the exhaust fan is on, otherwise, the exhaust fan is off. The test shows that the actual result is consistent with the preset. The stability of the data will be tested. The stability test is mainly aimed at the reliability test of communication and transmission between the sensor and the relay terminal. In this design, there are four temperature and humidity sensors for data collection, each collection cycle is 1 s. The total time of data collection is half an hour, and the results are shown in Table 2.

As can be seen from Table 2, a total of 1,800 times were collected within half an hour, and the transmission rate was about 97.6%. Data occasionally lost packets in the transmission process and the system stability was strong.

Table 2. Stability test

Status data acquisition node	Number of times sent	Number of times received	Transmission success rate/%
Workshop 1	1800	1752	97.3
Workshop 2	1800	1768	98.2
Machine room	1800	1747	97.1
Distribution room	1800	1759	97.7

6 Conclusions

This paper implements an intelligent exhaust system based on LORA communication. The system regards data acquisition as the underlying node, regards relay as hubs of information collection and judgment, and regards upper computer as the control interface of data processing and storage. The temperature and humidity of each node are remotely monitored and controlled through LORA communication. The communication protocol of this system greatly reduces the loss rate of data and ensures the stability of the system. This is very important for the factory to realize intelligence.

Acknowledgements. This paper is supported by the opening project of Nantong intelligent information technology joint research center (KFKT2017B05) and the first phase project of Jiangsu university brand specialty construction project (PPZY2015B135).

References

1. Tian, S.: Thermal control of the temperature control furnace and the upper computer design. In: International Society for Information and Engineering Proceedings of 2016 5th International Conference on Advanced Materials and Computer Science (ICAMCS 2016). International society for information and engineering, vol. 4 (2016)
2. Khandaker, M.R.A., Wong, K.-K.: Joint source and relay optimization for interference MIMO relay networks. EURASIP J. Adv. Signal Process. **2017**(1), 1–14 (2017)
3. Yu, X.: Research on intelligent exhaust device design of the breaker in the substation. China University of science and technology(Taiwan), Taiwan institute of construction engineering. In: Proceedings of the 2016 International Conference on Civil, Architecture and Environmental Engineering Vol. 1. China university of science and technology(Taiwan)、Taiwan institute of construction engineering, vol. 4 (2016)
4. Li, J., Meng, X., Song, W.: Intelligent temperature control system design based on single-chip microcomputer. J. Harbin Inst. Technol. **21**(03), 91–94 (2014)
5. Zhang, H.-f., Kang, W.: Design of the data acquisition system based on STM32. Procedia Comput. Sc. **17** (2013)
6. Guo, L., Liu, D., Zhang, X., Sun, X.: Design of intelligent detecting system based on wireless transmission. J. Zhejiang University SCIENCE A **8**(3), 490–494 (2007)

Waveform Design for Joint Radar-Communication Geosynchronous Broadcasting Satellite Systems

Liang Xu, Yicheng Jiang(✉), and Zhaofa Wang

Harbin Institute of Technology, 92 West Dazhi Street, Nangang District, Harbin 150001, China
Jiangyc@hit.edu.cn

Abstract. In recent years, the joint radar-communication geosynchronous (GEO) broadcasting satellite system has attracted more and more attention for its many advantages. For the joint system, a critical issue needs to be properly addressed is the waveform design. In this paper, we design a newly integrated waveform for the GEO satellite system, where the direct-sequence spread spectrum (DSSS) method is employed to obtain desired performance in both radar and communication. The simulations of bit error rate (BER) and auto-correlation function (ACF) validate the performance of the new waveform.

Keywords: Waveform design · Joint radar-communication · Direct-sequence spread spectrum

1 Introduction

The concept of joint radar-communication system, which means that the radar and communication functions can be integrated into a single system, was first proposed as the "multifunction RF" in [1]. The joint radar-communication geosynchronous (GEO) broadcasting satellite system has received much interest for its wide area coverage, high level of Equivalent Isotropic Radiated Power (EIRP), and simplicity of operation (not requiring tracking the satellite position) [2, 3].

The integrated waveform design is an essential step in the design of a joint radar-communication GEO broadcasting satellite system. In [4], the possibility of using the existing satellite waveforms directly for aerial target detection has been investigated. This research shows that the existing satellite waveform can be used to perform radar and communication functions simultaneously. However, the conventional satellite waveform is dedicated designed for communication, its limited bandwidth leads to the long coherent integration time and poor range resolution performance in radar operation. As the radar performance of the joint system is as important as its communication performance, the existing satellite waveform cannot be used as the integrated waveform directly. Consequently, a newly integrated waveform, which can be used to obtain desirable performance in both radar and communication, needs to be designed.

An ideal integrated waveform should have the following properties: (1) meets the dedicated satellite waveform standard (we use the Digital Video Broadcasting—Second

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 870–876, 2019.
https://doi.org/10.1007/978-981-13-6264-4_102

generation (DVB-S2) standard [5] in this paper) to ensure it can be transmitted by the satellite; (2) has high spectral efficiency because the bandwidth of the waveform is eventually limited by the satellite transponder [6]; (3) has good autocorrelation properties to obtain the good radar performance; (4) has good BER performance. In order to obtain the ideal integrated waveform which can be used for the joint radar-communication GEO broadcasting satellite system, the 32APSK modulation scheme and the DSSS method are employed in this paper. On one hand, high spectral efficiency can be obtained with the 32APSK modulation scheme. On the other hand, the BER performance and autocorrelation properties are improved by the DSSS method. The simulation results show that the new waveform has good performance.

2 Waveform Design

As the transmitted signal of the GEO broadcasting satellite, the integrated waveform should meet the DVB-S2 standard. Considering the high spectral efficiency allows waveform to have higher data rates within a limited bandwidth, according to the standard, the 32APSK modulation scheme is adopted in this paper for the high spectrum efficiency. However, the 32APSK waveform requires a relative high signal-to-noise ratio (SNR) to achieve the data transmission. As a consequence, the DSSS method is employed to (1) make waveform can perform communication function under low SNR conditions; (2) increase the bandwidth of the waveform to obtain better radar performance.

2.1 Modulation Scheme of the Integrated Waveform

Figure 1 shows the functional block diagram of the DVB-S2 waveform generation. The concatenation of Bose–Chaudhuri–Hocquenghem (BCH) outer codes and Low-Density Parity Check (LDPC) inner codes (rates 1/4, 1/3, 2/5, 1/2, 3/5, 2/3, 3/4, 4/5, 5/6, 8/9, 9/10) carry out the Forward Error Correction (FEC) Encoding. After the bit interleaving, the Quaternary Phase-Shift Keying (QPSK), 8PSK, 16APSK, and 32APSK modulation schemes shall be applied depending on the application area.

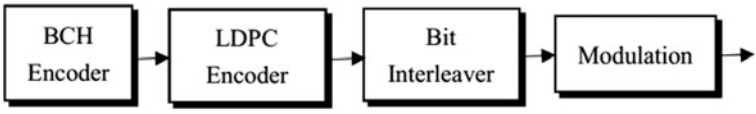


Fig. 1. Functional block diagram of the DVB-S2 waveform generation

The spectral efficiency and the ideal Es/No are employed to indicate the performance of the modulation schemes, where the ideal Es/No denotes the minimum SNR of the received signal that guarantees a quasi-zero BER transmission of the satellite system, Es is the signal energy in Joules, No is noise power spectral density. The low level of ideal Es/No means that the system can work under low SNR conditions, in other words, under the same SNR, the communication performance of the waveforms

degrades with the increase of the order modulation scheme. Table 1 shows the spectral efficiencies and the ideal Es/No of the four kinds of modulation schemes with a fixed inner code rate 3/4. As we can see, both the spectral efficiency and the ideal Es/No of the corresponding waveform increase as the order of the modulation scheme increases, that is to say, the high-order modulation schemes with a high spectral efficiency requires a relatively high SNR to guarantee its data transmission performance.

Table 1. DVB-S2 standard

Modulation scheme	Spectral efficiency	Ideal Es/No (dB)
QPSK3/4	1.487473	4.03
8PSK3/4	2.228124	7.91
16APSK3/4	2.966728	10.21
32APSK3/4	3.703295	12.73

It can be seen that there is a trade-off between the spectral efficiency and communication performance. As mentioned before, an ideal integrated waveform is expected to have high spectral efficiency and good communication performance at the same time. To obtain this kind of waveform, two methods can be employed: (1) adopt the low-order modulation scheme and improve the spectral efficiency; (2) adopt the high-order modulation scheme and improve the communication performance. Apparently, the spectral efficiency is only determined by the fixed coding and modulation scheme of the corresponding waveform while the communication performance of the waveform can be improved by existing methods. Consequently, we like to choose the second method to design the new waveform, namely, adopt the high-order modulation scheme and find a method to improve its communication performance. As the 32APSK modulation scheme has the highest spectral efficiency, we employ this scheme to modulate the newly integrated waveform

2.2 DSSS Method

Due to the long distance between the satellite and the receiver on the ground, the satellite system works under low SNR conditions generally. In this situation, using the 32APSK waveform directly requires a long coherent time for target detection. And also, the limited bandwidth of the 32APSK waveform leads to the poor range resolution performance. Moreover, the 32APSK waveform requires a relatively high SNR to ensure its communication performance.

DSSS is a modulation method applied to digital signals. It increases the signal bandwidth to a value much larger than needed to transmit the information. As it is well known, the DSSS method takes advantages of the correlation properties of spreading codes to enable both radar and communication systems to work under the low SNR condition [7]. Therefore, we use the DSSS method on the 32APSK waveform to obtain the better radar and communication performance of the system.

The diagram of the proposed waveform generation is shown in Fig. 2. The waveform at the output of 32APSK modulator can be expressed as

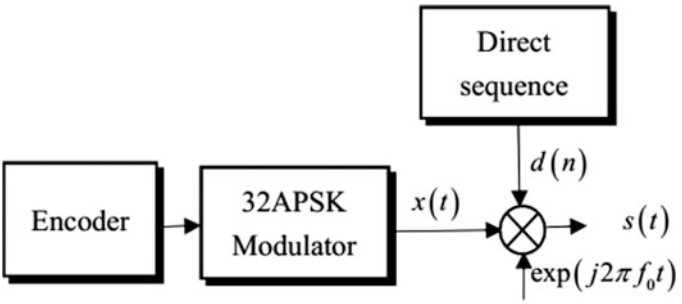


Fig. 2. Waveform generation

$$x(t) = \sum_{i=0}^{N_{sym}-1} a(i) rect\left(\frac{t - iT_{sym}}{T_{sym}}\right) \tag{1}$$

where N_{sym} is the number of modulation symbols, $a(i)$ denotes the discrete amplitude and phase state of 32APSK modulation scheme, $rect\left(t/T_{sym}\right)$ is a rectangular window of duration T_{sym} and T_{sym} denotes the symbol duration.

Let $d(n)$ denotes a direct sequence, where $d(n) \in \{1, -1\}$. Then the final waveform $s(t)$ can be expressed as

$$\begin{aligned} s(t) &= \exp(j2\pi f_0 t) \\ &\times \sum_{i=0}^{N_{sym}-1} \sum_{n=1}^{L} d(n) a(i) rect\left(\frac{t - nT_{chip}}{T_{chip}}\right) rect\left(\frac{t - iT_{sym}}{T_{sym}}\right) \end{aligned} \tag{2}$$

where f_0 is the carrier frequency, L denotes the length of the sequence and we define it as the spread factor, T_{chip} denotes the chip duration with $T_{chip} = T_{sym}/L$.

3 Simulation Results

In this section, we investigate the BER performance and the autocorrelation function (ACF) of the integrated waveform to validate its performance of communication and radar respectively.

The parameters used in the simulation are given in Table 2. We assume that the spread spectrum operation just affects the chip rate of the waveform, and the data rate of the waveform at the output of the modulator equals to 1Mbit/s. As the bandwidth of the proposed waveform is eventually limited by the satellite transponder, we assume

Table 2. Parameters in simulation

Parameter	Value
Carrier frequency	11.92 GHz
Data rate (at the output of modulator)	1 Mbit/s
Transponder bandwidth	36 MHz
Modulation scheme	32APSK
Direct sequence	M-sequence
Spreading factor L	Variable (7,15,31,63,127)

the bandwidth of the transponder equals to 36 MHz. The M-sequence is used for the spread spectrum.

The BER performance of the waveforms under different SNRs is investigated, where the channel is assumed as the additive white Gaussian noise channel. The BER results of the waveforms are compared in Fig. 3, where the 32APSK waveform denotes the waveform without spread spectrum, the spectrum of integrated waveform is spread based on the 32APSK waveform, where L denotes the spread factors of the integrated waveforms. It can be seen that, (1) the BER performance is significantly improved after employing the DSSS method, that is to say, the proposed waveform make the system allows for operation under the condition of low SNR; (2) with the increase of the spread factor, the BER performance of integrated waveform is improving. However, the bandwidth of the waveform is eventually restricted by the bandwidth of transponder, the spread factor cannot increase indefinitely, in other words, there is a trade-off between the BER performance and the data rate when the bandwidth of the transponder is fully used by the integrated waveform. The spread factor and data rate can be set depending on the practical scenarios.

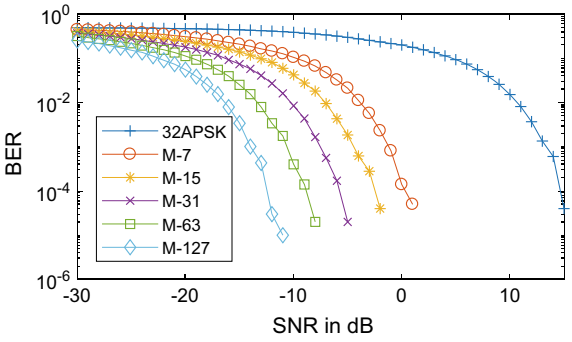


Fig. 3. BER performance of the waveforms

To investigate the performance of the new waveform in radar operation, the ACF of the waveform is studied. Figure 4 shows the comparison of the ACF results of the 32APSK waveform and the integrated waveform with the spread factor $L = 7$.

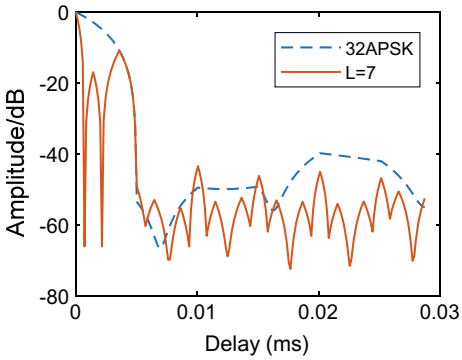


Fig. 4. ACF results of the 32APSK waveform and the integrated waveform with $L = 7$

As shown in Fig. 4, the main lobe of the ACF of the integrated waveform is narrower than that of the 32APSK waveform. As well known, the main lobe width of the ACF indicates the range resolution in radar operation, a narrow main lobe means high range resolution. Apparently, after using the DSSS method, the bandwidth of the integrated waveform is larger than that of the 32APSK waveform, which benefits the range resolution in target detection consequently.

As mentioned before, with a fixed data rate, the bandwidth of the integrated waveform is proportional to the spread factor. A large spread factor leads to the better radar performance of the new waveform. As the bandwidth of the waveform is restricted by the satellite transponder, we can choose a suitable spread factor for the proposed waveform to make full use of the transponder in practical applications.

4 Conclusions

In this paper, we design a newly integrated waveform for the joint radar-communication GEO broadcasting satellite system. According to the DVB-S2 standard, the integrated waveform employs the 32APSK modulation scheme to obtain the high spectral efficiency. The DSSS method is employed for improving the radar performance of the integrated waveform, and decreasing the SNR requirement of communication. The simulation results validate that the integrated waveform has improved BER and ACF performance, which means the waveform can be used to obtain good performance in communication and radar detection.

Acknowledgements. This paper is supported by the China Scholarship Council and the National Natural Science Foundation of China (Grant No. G61271342).

References

1. Tavik, G.C., Hilterbrick, C.L., Evins, J.B., et al.: The advanced multifunction RF concept. IEEE Trans. Microw. Theory Tech. **53**(3), 1009–1020 (2005)
2. Cristallini, D., Caruso, M., Falcone, P., et al.: Space-based passive radar enabled by the new generation of geostationary broadcast satellites. In: IEEE Aerospace Conference, Big Sky, USA, March 2010, pp. 1–11 (2010)
3. Zhang, D., Qiu, X., Hu, D., et al.: A subspace-based channel calibration algorithm for geosynchronous satellite-airborne bistatic multi-channel radars. IET Radar Sonar Navig. **8**(9), 1008–1017 (2014)
4. Wei, J., Feng, Y., Baohua, Z.: Analysis on guidance performance based on geostationary broadcast satellite. In: IET International Radar Conference, Hangzhou, China, pp. 1–4 (2015)
5. ETSI. En 302 307 v1.2.1: Digital video broadcasting (dvb); second generation framing structure, channel coding and modulation systems for broadcasting, interactive services, news gathering and other broadband satellite applications (dvb-s2) (2009–08)
6. Fred, N.E.: Radar Design Principles, 2nd edn. McGraw-Hill, New York (1991)
7. Quyen, N.X., Long, N.H., Van Yem, V.: An OFDM-based chaotic DSSS communication system with M-PSK modulation. In: Fifth International Conference on Communications and Electronics (ICCE), pp. 106–111. IEEE (2014)

Power Control of Full-Duplex D2D Communication Underlying Cellular Networks

Yuyun Xia[1], Shaolin Ma[1], and Liang Han[1,2(✉)]

[1] College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin 300387, China
hanliang_tjnu@163.com
[2] Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China

Abstract. Full-duplex device-to-device communication (D2D) communications underlying cellular networks can share the resources of cellular users and thereby can improve spectrum efficiency and overall system throughput. This paper presents a low-complex power control algorithm to solve the problem that a pair of full-duplex D2D users reuse the uplink channel resources of a single cellular user, and full-duplex D2D users interfere with the cellular user. The algorithm maximizes the throughput while ensuring the quality of service (QoS) of the D2D users and cellular user. The simulation results show that the algorithm can improve the throughput and the throughput depends on the QoS requirements and the self-interference cancelation ability.

Keywords: Full-duplex · D2D · Power control

1 Introduction

At present, wireless communication is gradually developing in the direction of network convergence. However, the non-renewability and low utilization of wireless spectrum resources have been the bottleneck in the development of wireless communication networks. As a key technology of 5G, device-to-device (D2D) communication allows nearby users to communicate directly without a base station, which can increase spectrum efficiency and transmission rate, reduce transmission delay and the terminal's power consumption and reduce the load of the base station. As another key technology, full duplex allows a device to transmit and receive signals at the same time and at the same frequency. Therefore, combining these two technologies will further increase the spectrum efficiency of cellular networks and the transmission rate. However, since the co-channel interference and residual self-interference exist, power control for full-duplex D2D communication underlying cellular networks has great research potential and value.

Reference [1] analyzed the SINR statistics of all cellular users (CUs) and D2D users to obtain the optimal transmit power of the D2D device. Reference [2] proposed a scheme for dynamically adjusting the transmit power of D2D users based on the real-time measurement information of the channel. Reference [3] investigated D2D power control to maximize total system throughput under total power constraint. However, the

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 877–885, 2019.
https://doi.org/10.1007/978-981-13-6264-4_103

above papers considered the power control technology in the half-duplex mode. The power control algorithm mentioned in [4] is a binary power control method that sets a maximum transmit power for the user. If the D2D user uses the resources of the cellular user, then its transmitting power is the maximum transmit power minus a threshold. This kind of power control algorithm is very simple, however, this power control algorithm is too granular, it does not take into account the position and distribution factors between cellular users and D2D users.

2 System Model

As shown in Fig. 1, in order to improve the efficiency of wireless spectrum resources and reduce the interference caused by multiplexing, this document assumes the uplink channel resources of a single cellular user. Only one D2D can be used to multiplex users, and a single pair of D2D users can only reuse the uplink channel resources of one cellular user.

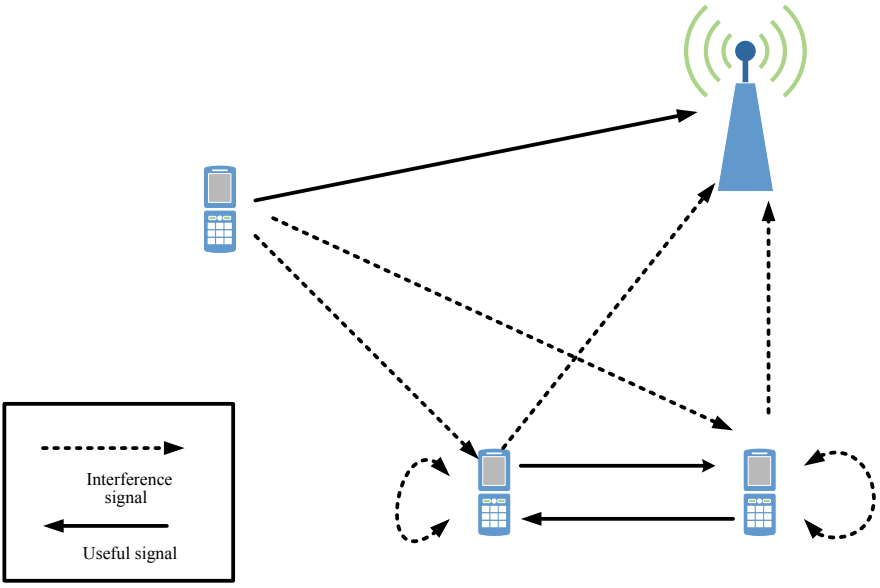


Fig. 1. D2D users use the uplink channel scenario

The base station can manage the communication of this communication system, including D2D discovery, connection, and power control and channel allocation. The process of optimal power control is that after the base station measures the interference of the received D2D user, it coordinates the transmit power of the cellular user and the D2D user through the control channel to meet the QoS requirements of the cellular user

and the D2D user. The signal-to-interference-plus-noise ratio of the cellular link, D2D users 1 and D2D users 2 are given as

$$\xi_c = \frac{P_c g_{CB}}{\sigma_n^2 + P_d g_{B1} + P_d g_{B2}} \geq \xi_{c,\min} \tag{1}$$

$$\xi_{d1} = \frac{P_d g_{12}}{\sigma_n^2 + P_c g_{c2} + \gamma P_d} \geq \xi_{d,\min} \tag{2}$$

$$\xi_{d2} = \frac{P_d g_{12}}{\sigma_n^2 + P_c g_{c1} + \gamma P_d} \geq \xi_{d,\min} \tag{3}$$

where P_c and P_d denote the transmit power of the cellular user and the D2D user, respectively, g_{CB}, g_{B1}, g_{B2}, g_{c2} and g_{c1} denote the channel gain of the cellular user to the base station, the D2D users 1 to the base station, the D2D users 2 to the base station, the cellular user to the D2D users 1, and the cellular user to D2D user 2, respectively. σ_n^2 stands for Gaussian white noise power, γ stands for self-interference cancelation factor.

3 Power Control Algorithm

For full-duplex transmission mode of D2D communication, multiplexing the CU's uplink channel resources can effectively improve resource reuse gain and link gain, and further improve the system's spectrum efficiency and network throughput. When a D2D link is established, it is necessary to satisfy both the minimum SINR requirement and the smaller interference to the multiplexed CU. Therefore, when the same wireless channel resources serve the same CU user and FD-D2D pair user at the same time, under the premise of guaranteeing the CU user and the D2D minimum SINR requirement, the D2D is solved by maximizing the throughput of the entire system link. The optimal transmit power of full-duplex D2D can be modeled as follows:

$$(P_c^*, P_d^*) = \arg\max_{(p_c, p_d)} \{\log_2(1+\xi_c) + \log_2(1+\xi_{d1}) + \log_2(1+\xi_{d2})\} \tag{4}$$

$$\text{s.t. } \xi_c = \frac{P_c g_{CB}}{\sigma_n^2 + P_d g_{B1} + P_d g_{B2}} \geq \xi_{c,\min} \tag{5}$$

$$\xi_{d1} = \frac{P_d g_{12}}{\sigma_n^2 + P_c g_{c2} + \gamma P_d} \geq \xi_{d,\min} \tag{6}$$

$$\xi_{d2} = \frac{P_d g_{12}}{\sigma_n^2 + P_c g_{c1} + \gamma P_d} \geq \xi_{d,\min} \tag{7}$$

$$P_c \leq P_{c,\max}, P_d \leq P_{d,\max} \tag{8}$$

Among them, (4) is the optimization goal, and (5)–(8) are constraints. The optimization variables are the transmit power of the D2D user and P_d the transmit power of the cellular user. $\xi_{c,\min}$, $\xi_{d,\min}$ the minimum signal-to-interference-plus-noise ratio for cellular users, D2D users 1 and 2, and $P_{c,\max}$, $P_{d,\max}$ the maximum transmit power for cellular users and D2D users, respectively.

For the above-mentioned optimization problem, it is difficult to directly obtain a global optimal solution. Therefore, by analyzing the objective function of the optimization model and the characteristics of the feasible restricted area boundary, an algorithm with low computational complexity is proposed. Through this algorithm, the optimal solution to the problem can be found.

The transmit power limitation areas of cellular users and D2D users are determined by Eqs. (5)–(7) are assumed to be represented by lines l_c, l_1, and l_2, respectively. This paper assumes that the D2D user's transmit power is equal. And the minimum signal-to-interference-plus-noise ratio of the two D2D users is the same.

l_c expression:

$$P_d = \frac{g_{CB}}{\xi_{c,\min}(g_{B1} + g_{B2})} P_c - \frac{\sigma_n^2}{g_{B1} + g_{B2}} \tag{9}$$

l_1 expression:

$$P_d = \frac{\xi_{d,\min} g_{C2}}{g_{12} - \xi_{d,\min}\gamma} P_c + \frac{\xi_{d,\min}\sigma_n^2}{g_{12} - \xi_{d,\min}\gamma} \tag{10}$$

l_2 expression:

$$P_d = \frac{\xi_{d,\min} g_{C1}}{g_{12} - \xi_{d,\min}\gamma} P_c + \frac{\xi_{d,\min}\sigma_n^2}{g_{12} - \xi_{d,\min}\gamma} \tag{11}$$

It can be seen that if the slope of the straight line l_c is smaller than the slope of the straight line l_1, the two constraints do not have a common solution region. At this time, the optimization problem has no solution. Therefore, to solve the optimization problem, the slope of the line l_c must be greater than the slope of the line l_1:

$$\frac{g_{CB}}{\xi_{c,\min}(g_{B1} + g_{B2})} > \frac{\xi_{d,\min} g_{C2}}{g_{12} - \xi_{d,\min}\gamma} > 0 \tag{12}$$

In order to make the original problem bounded, it must also require that the intersection of l_c and l_1, l_2 lies in a rectangular box consisting of $(P_{c,\max}, P_{d,\max})$. The constraints have three types of possible restricted areas, as shown by the shaded areas in Fig. 2, for the various possibilities of the straight line l_2, and therefore there are also multiple situations in each class.

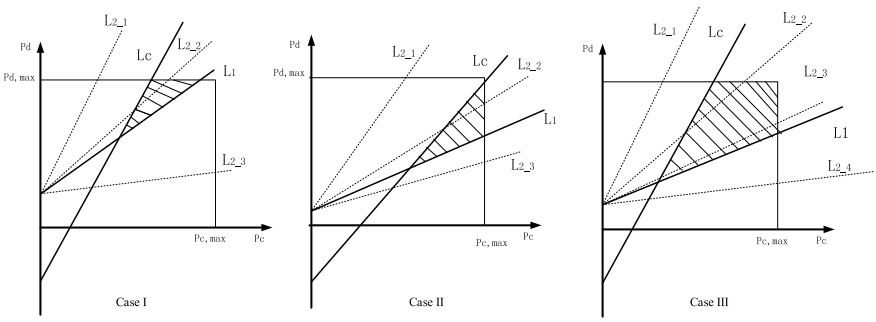


Fig. 2. Optimal potential power point of the system

It can be seen that the points in all shaded regions in the three cases of graphs satisfy the condition. Therefore, the optimal power point can be found in these shaded regions.

$$f(P_c, P_d) \triangleq \log_2(1 + \frac{P_c g_{CB}}{\sigma_n^2 + P_d g_{B1} + P_d g_{B2}}) + \log_2(1 + \frac{P_d g_{12}}{\sigma_n^2 + P_c g_{c2} + \gamma P_d}) + \log_2(1 + \frac{P_d g_{12}}{\sigma_n^2 + P_c g_{c1} + \gamma P_d}) \tag{13}$$

It is easy to see

$$f(\lambda P_c, \lambda P_d) \triangleq \log_2(1 + \frac{P_c g_{CB}}{\frac{\sigma_n^2}{\lambda} + P_d g_{B1} + P_d g_{B2}}) + \log_2(1 + \frac{P_d g_{12}}{\frac{\sigma_n^2}{\lambda} + P_c g_{c2} + \gamma P_d}) + \log_2(1 + \frac{P_d g_{12}}{\frac{\sigma_n^2}{\lambda} + P_c g_{c1} + \gamma P_d}) \tag{14}$$

when λ is greater than 1, it is a strictly increasing function. Therefore, for any point in the feasible area (P_c, P_d), there must be another point $(\lambda P_c, \lambda P_d)$, when $\lambda > 1$, it makes $f(\lambda P_c, \lambda P_d) > f(P_c, P_d)$. Therefore, you can gradually increase the value of any point in the feasible region of the shadow to increase the value of the optimization goal. At this point, there must be a point to reach the boundary of the feasible region. Therefore, when the optimization target obtains the maximum value, the optimization variable must be located at the boundary.

Case I: The limit power point $(P_{c,\max}, P_{d,\max})$ is on the right side of the straight line l_1. When $\frac{P_{d,\max} g_{12}}{\sigma_n^2 + P_{c,\max} g_{c2} + \gamma P_{d,\max}} < \xi_{d,\min}$, at this time, due to the uncertainty of the D2D user's minimum signal-to-noise ratio $\xi_{d,\min}$, l_2 has three conditions: l_{2_1}, l_{2_2}, and l_{2_3}.

The solution to the optimal problem at this time can be described as

$$
\begin{aligned}
(P_c^*, P_d^*) &= \underset{(P_c,P_d)\in\phi_1}{\arg\max} f(P_c, P_d) \\
\phi_1 &= \left\{ \left(P_1, P_{d,\max}\right), \left(P_2, P_{d,\max}\right), \left(P_3, P_{d,\max}\right) \right\} \\
P_1 &= \frac{P_{d,\max}\left(g_{12} + \xi_{d,\min}\gamma\right) - \xi_{d,\min}\sigma_n^2}{\xi_{d,\min} g_{C1}} \\
P_2 &= \frac{P_{d,\max}\xi_{c,\min}\left(g_{B1} + g_{B2}\right)}{g_{CB}} + \frac{\xi_{d,\min}\sigma_n^2}{g_{CB}} \\
P_3 &= \frac{P_{d,\max}\left(g_{12} + \xi_{d,\min}\gamma\right) - \xi_{d,\min}\sigma_n^2}{\xi_{d,\min} g_{C2}}
\end{aligned}
\tag{15}
$$

Case II: When the limit power point $(P_{c,\max}, P_{d,\max})$ is on the left side of the straight line l_c, that is when ($\frac{P_{c,\max} g_{CB}}{\sigma_n^2 + P_{d,\max} g_{B1} + P_{d,\max} g_{B2}} < \xi_{c,\min}$ and $\frac{P_{d,\max} g_{12}}{\sigma_n^2 + P_{c,\max} g_{c2} + \gamma P_{d,\max}} > \xi_{d,\min}$) At this time, due to the uncertainty of the D2D user's minimum S/N ratio $\xi_{d,\min}$, l_2 still has three cases: l_{2_1}, l_{2_2}, and l_{2_3}.

The solution to the optimal problem at this time can be described as

$$
\begin{aligned}
(P_c^*, P_d^*) &= \underset{(P_c,P_d)\in\phi_2}{\arg\max} f(P_c, P_d) \\
\phi_2 &= \left\{ \left(P_{c,\max}, P_4\right), \left(P_{c,\max}, P_5\right), \left(P_{c,\max}, P_6\right) \right\} \\
P_4 &= \frac{g_{CB}}{\xi_{c,\min}\left(g_{B1} + g_{B2}\right)} P_{c,\max} - \frac{\sigma_n^2}{g_{B1} + g_{B2}} \\
P_5 &= \frac{\xi_{d,\min} g_{C2}}{g_{12} + \xi_{d,\min}\gamma} P_{c,\max} - \frac{\xi_{d,\min}\sigma_n^2}{g_{12} + \xi_{d,\min}\gamma} \\
P_6 &= \frac{\xi_{d,\min} g_{C1}}{g_{12} + \xi_{d,\min}\gamma} P_{c,\max} - \frac{\xi_{d,\min}\sigma_n^2}{g_{12} + \xi_{d,\min}\gamma}
\end{aligned}
\tag{16}
$$

Case III: When the limit power point $(P_{c,\max}, P_{d,\max})$ is on the right side of the straight line l_c and is on the left side of l_1, that is when ($\frac{P_{c,\max} g_{CB}}{\sigma_n^2 + P_{d,\max} g_{B1} + P_{d,\max} g_{B2}} > \xi_{c,\min}$), at this time, due to the uncertainty of the D2D user's minimum signal to noise ratio $\xi_{d,\min}$, l_2 has four conditions: l_{2_1}, l_{2_2}, l_{2_3},l_{2_4}.

The solution to the optimal problem at this time can be described as

$$
\begin{aligned}
(P_c^*, P_d^*) &= \underset{(P_c,P_d)\in\phi_3}{\arg\max} f(P_c, P_d) \\
\phi_3 &= \left\{ \left(P_{c,\max}, P_7\right), \left(P_{c,\max}, P_8\right), \left(P_{c,\max}, P_{c,\max}\right), \left(P_9, P_{d,\max}\right), \left(P_{10}, P_{d,\max}\right) \right\} \\
P_7 &= P_6, P_8 = P_5, P_9 = P_1, P_{10} = P_2
\end{aligned}
\tag{17}
$$

Summary: According to the above three conditions, it is possible to find the optimal power solution for D2D users and cellular users, and then the overall system throughput is maximized.

4 Simulation Results

The simulation scenario is that within a single cell, a single pair of D2D users multiplexes a single cellular user resource. The distance d1 between the cellular user and the base station is a fixed 500 m. The D2D user moves on the horizontal line d2 between the cellular user and the base station user. The simulation channel considers both large-scale fading and small-scale fading. The simulation scenario is shown in Fig. 3.

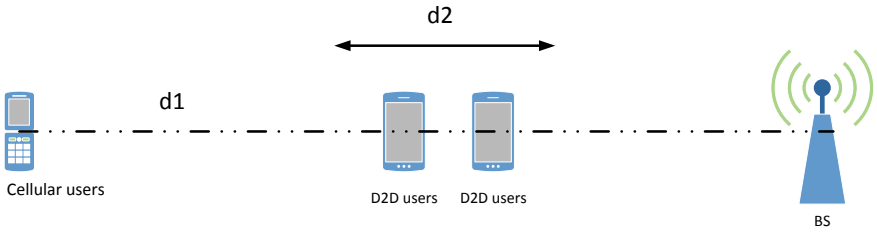


Fig. 3. Simulation scenario

The maximum distance between D2D users is 50 m. D2D users use the same frequency full duplex working mode while multiplexing uplink resources. The specific simulation parameters are as follows (Table 1):

Table 1. Simulation parameter settings

Simulation parameters	Value
Maximum distance between D2D users	80 m
Distance between cellular users and base stations	500 m
Cellular user's maximum transmit power P_C^{MAX}	25 dBm
Maximum transmit power of D2D users P_{D2D}^{MAX}	24 dBm
Noise power σ_N^2	−174 dBm
Path loss factor α	4
Minimum SNR of cellular users $\xi_{c,\min}$	8–15 dB
D2D user's minimum signal-to-noise ratio $\xi_{d,\min}$	8 dB

First: In order to compare the performance of the proposed algorithm, the simulation will be compared with the previously mentioned binary power control algorithm, and in order to compare the advantages of a full-duplex communication system, the performance of the half-duplex link is compared. Figure 4 compares the performance of the two power control algorithms in full-duplex mode and compares them with the throughput of a half-duplex link. The minimum SINR requirement for cellular and D2D users is 10 dB and 8 dB, and the self-interference cancelation capability.

It is 10^{-12}. It can be seen from Fig. 4 that for both power control algorithms, the throughput of the system will decrease as the distance between D2D users increases. The reason is that as the distance d of the D2D user increases, the signal power gain between the D2D user's decreases, and in order to meet their own minimum signal-to-noise ratio requirement, the D2D user's own transmit power will be increased, and thus it will be recovered. The use of cellular users creates more interference.

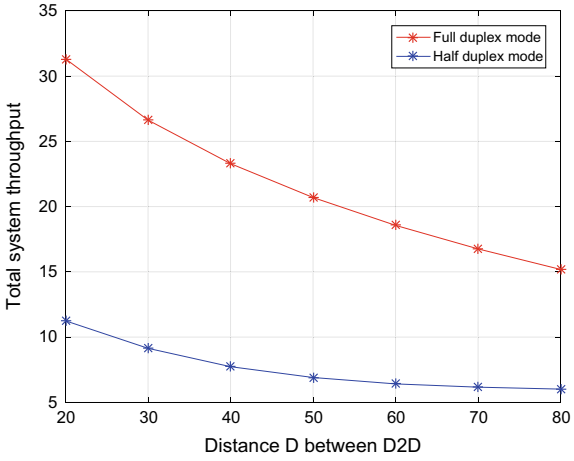


Fig. 4 Comparison of different algorithms

Figure 5 examines the effect on overall system performance for different self-interference cancelations. Among them, the minimum signal to interference and noise

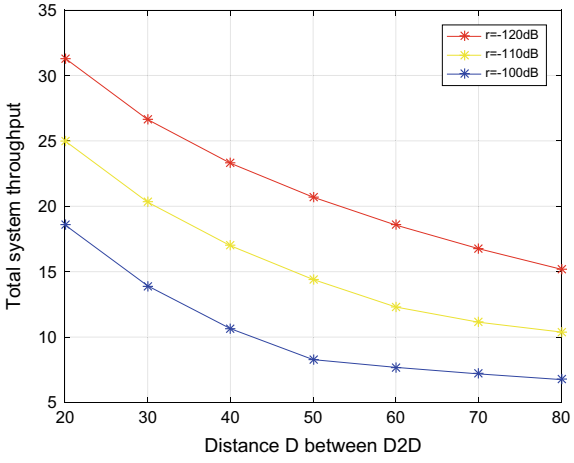


Fig. 5. Effect of self-interference cancelation on the system

ratio of cellular users and D2D users is 10 dB and 8 dB. Since more and more self-interference is eliminated, the signal-to-interference ratio of D2D links becomes larger and larger, and there is no significant effect on the S/N ratio of cellular users. Therefore, the performance of the D2D link is becoming more and more superior, resulting in superior performance of the entire link.

5 Conclusion

In this paper, a low-complexity power control algorithm is proposed to solve the interference problem of simultaneous co-channel full-duplex communication systems and ultimately maximize the overall throughput of the system. The simulation results show that by introducing a D2D communication system with simultaneous frequency and full duplex, the total cell throughput is greatly improved under the guarantee of the service quality of the cellular user of the original cellular system and the service quality of the D2D user.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (61701345), National Natural Science Foundation of Tianjin (18JCZDJC31900), and Tianjin Education Commission Scientific Research Plan (2017KJ121).

References

1. Yu, C.H., Tirkkonen, O., Doppler, K.: On the performance of device-to-device underlay communication with simple power control. In: IEEE International Conference on VTC, pp. 1–5 (2009)
2. Jaheon, G., Sueng, J.B., Bumgon, C.: Dynamic power control mechanism for interference coordination of device-to-device communication in cellular networks. In: IEEE International Conference on ICUFN, pp. 71–75 (2011)
3. Tirkkonen, O., Doppler, K., Ribeiro, C.: Power optimization of device-to-device communication underlaying cellular communication. In: IEEE International Conference on ICC, pp. 1–5 (2009)
4. Ali, S., Rajatheva, N., Latva-aho, M.: Effect of interference of full-duplex transmissions in underlay device-to-device communication. In: IEEE 14th Canadian Workshop on Information Theory (CWIT), Toronto, Canada, pp. 54–57 (2015)

The Satellite-Based M2M Services for Smart Grids: Case Study

Ning Jiang[1,2], Chengjun Yang[1], Xiaobo Guo[2], Zhijian Zhang[1], and Jian Wang[1(✉)]

[1] Nanjing University, Nanjing, China
{jiangning1022,1975817668}@qq.com,
{njuzzj,wangjnju}@nju.edu.cn
[2] Science and Technology on Information Transmission and Dissemination in Communication Networks Laboratory, The 54th Institute of CETC, Shijiazhuang, Hebei, China
guoxiaobo@foxmail.com

Abstract. The smart grid (SG) is an intelligent evolution of the electricity network that depends on highly reliable and secure connectivity between various infrastructure components from utilities, in order to meet the requirements of smart metering communication. Machine-to-machine (M2M) technology is designed for automatic equipment when it exchanges data, which is suitable for SG; Wireless communication is a main mode of M2M, and satellite communication provides support for the coverage of SG communication network. In this paper, an overview of the satellite-based M2M service for the smart grid is provided, and a case system is discussed with Beidou-based M2M service for smart grid.

Keywords: Smart grid · M2M · Satellite

1 Introduction

At present, the communication infrastructure of smart grid includes wired and wireless. Wireline-based solutions have available and developing, but wireless communication technologies will become increasingly important, especially in rural environments, scenes that need long-distance communication. All of that can greatly benefit from wireless or satellite-based system for the smart grid. Wireless automatic meter reading is of particular interest to utilities for both urban and suburban environments [1].

M2M has been applied more and more in smart grid [2–5]. M2M communication happens between machines, this practice often ignores direct manual intervention. Smart grid provides conditions for devices that require large numbers of automated operations.

Advanced Metering Infrastructure (AMI) is a significant component of the SG, designed to be Demand Response-enabled [6], the two-way communication between power companies and consumers enable consumers to monitor and adjust the use of energy as well. Figure 1 shows a schematic diagram of typical AMI architecture, it contains four components: Meter Database Management system (MDMS),

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 886–890, 2019.
https://doi.org/10.1007/978-981-13-6264-4_104

Neighborhood Area Network (NAN), Wide Area Network (WAN), and Home Area Network (HAN). Increasingly, companies are planning to launch intelligent metering services by incorporating M2M devices in the meters.

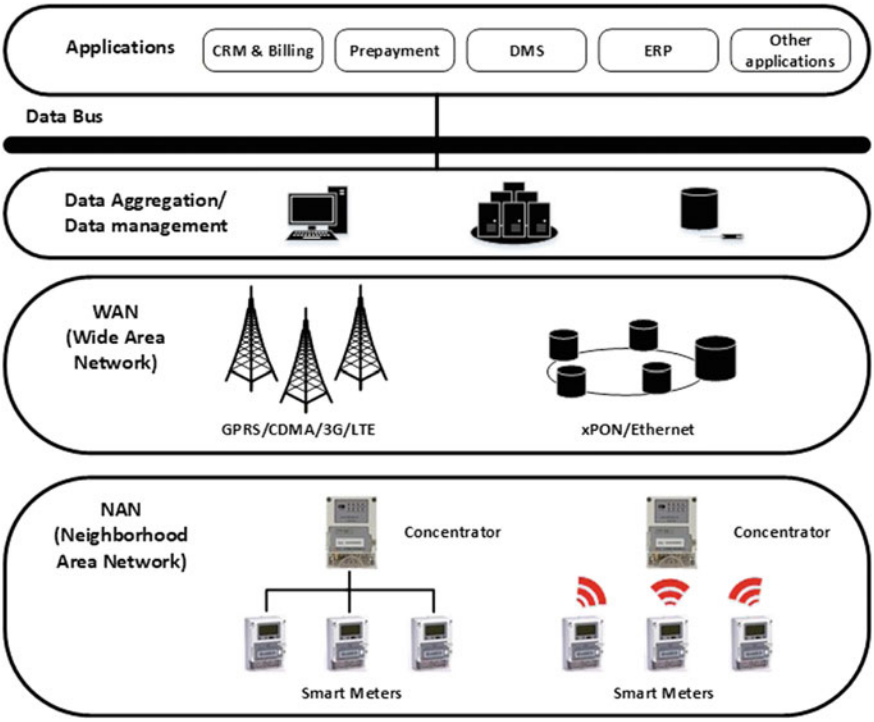


Fig. 1 Typical AMI architecture

In our proposed model, we focus on a case study of the satellite-based M2M service for smart grid: Beidou-based M2M service for the smart grid. Section 2 introduces the development and application of the satellite-based M2M service; Sect. 3 discusses the present situation of Beidou-based M2M service for smart grid, and our model.

2 The Satellite-Based M2M Service

The satellite-based M2M service provided connectivity over large areas. Since terrestrial networks may not service all locations on earth, satellite operators, low earth orbit satellites such as Iridium, Mid orbit satellites such as Inmarsat, and Beidou offer global connectivity services that can extend M2M communication to almost 100% of the earth. Hence, the delivery of relatively small quantities of data in almost all cases is supported by the satellite-based M2M service [1].

Over the past decade, M2M communication through satellite communications, especially with the development of satellite networks, have played an important role in many fields. A typical application is the sensor network [8], M2M sensor devices are grouped into clusters and communicate with satellite gateways. The characteristics of M2M communication system are a large number of equipment, low data rate, and large coverage area, which fits the demand of the smart grid. The satellite-based M2M service in the smart grid is similar to the satellite-based M2M sensor network, and the smart meter (concentrator) is grouped into clusters in the local and communicate with satellite gateway. It is a workable scheme to collect electrical information from the Beidou short message. At the end of 2012, the system was formally put into operation in the three functions in the Asia Pacific region, including Navigation and Position, Time Giving, and Short Message. The Beidou satellite communication system has a wide range of coverage, no communication blind district and enciphering of message [7].

Therefore, we take the application of M2M service in smart grid based on Beidou satellite as a case.

3 Case: The Beidou-Based M2M Service

3.1 Demands for SG and Characteristics of BD

Beidou satellite communication has its limitations, so it needs to be perfect for the relevant measures to meet the needs of smart grid. The characteristics of Beidou (BD) satellite communication and the demand for smart grid communication are shown in Table 1.

Table 1 Demands for SG and characteristics of BD

	Demand for SG	Characteristics of BD
Two-way channels	Yes	Yes
Widespread coverage	Yes	Yes
Adequate quality of service	Yes	No
Scalability	Yes	Yes
Confidentiality	Yes	Yes
Channel capacity	More than 100(Almost)	78 (civil)
Response time	Less than 20 s	60 s (civil)
Communication protocol	The Power Communication Protocol	Beidou 4.0 Protocol

From the table, Beidou satellite communication system cannot meet all the needs of smart grid communication, we need to make some changes.

3.2 System Architecture

We need to make some improvements when Beidou satellite communication is added to WAN in the AMI system. As seen in Fig. 2, the addition of a gateway between

MDMS and WAN, between WAN and NAN, can ensure that the original network can be changed as little as possible, and solve the existing problems properly. The ARQ coding scheme can be used to solve the problem that the Beidou satellite communication cannot guarantee reliable transmission in the transmission of power data into the communication network; The long message segmentation method can be used for the problem of the lack of the capacity of the Beidou satellite communication, it breaks data into packets in this end and combines incoming message into a single message in other end; On the response speed problem, we use gateway to response the end, the

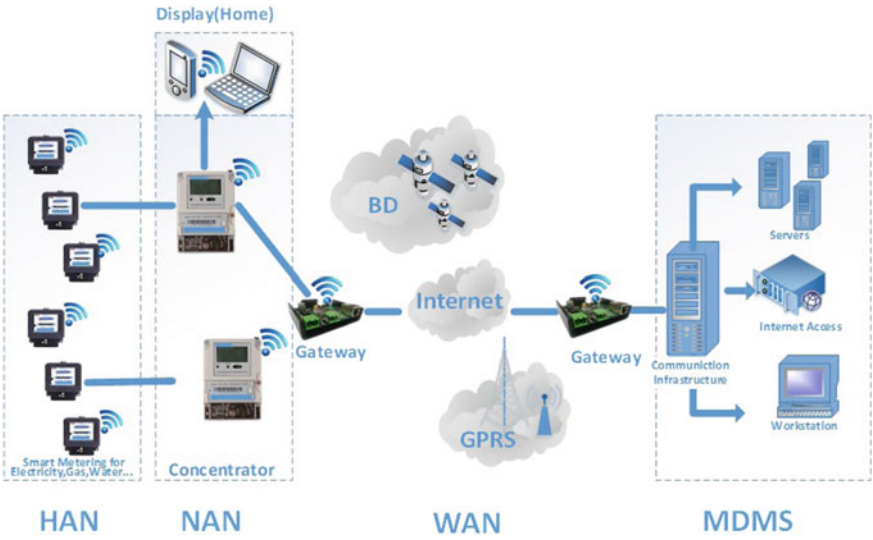


Fig. 2 Satellite-based M2M service for smart grid

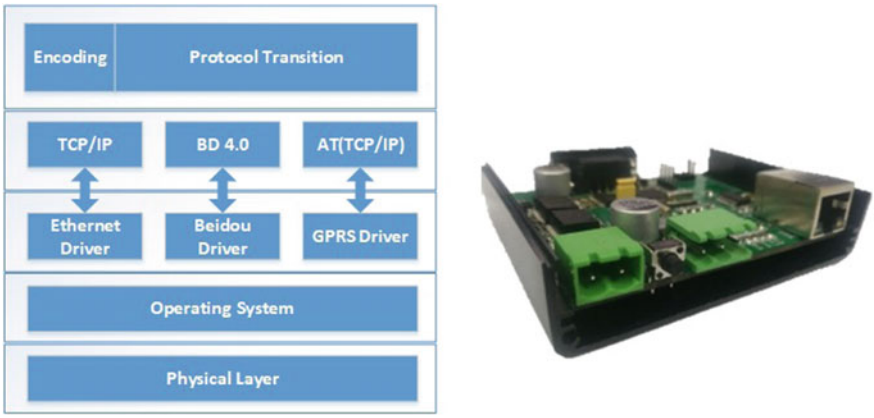


Fig. 3 Satellite gateway

new response strategy is specified between the gateway, the gateway converts the protocol as well.

But even so, the time needed to send a longer message is still too long, so the longer message can be compressed first, such as LZ77, which can reduce the length of the message to be sent, which will reduce the time cost.

The gateway structure is shown in Fig. 3. The physical layer provides physical access for connection; the operating system loads the related drivers, following the corresponding protocol; encoding (long message enhancement, ARQ coding, LZ77 coding, etc.) and protocol conversion happen in the application layer.

4 Conclusion

We summarize the effect of M2M and satellite communication in smart grid, which is demand-focused communication network. The satellite-based M2M service has a good prospect in many fields. More and more studies have begun to be attempted in smart grid for characteristics of this system. The Beidou satellite communication is running well in the Asia Pacific region and has been widely used. Therefore, the Beidou-based M2M service is taken as a case to solve the problems, which exists in the fusion process of the satellite-based M2M service for smart grid.

Acknowledgements. This work was supported by State Key Laboratory of Smart Grid Protection and Control of NARI Group Corporation, and the Postgraduate Research and Practice Innovation Program of Jiangsu Province (SJCX18_0004).

References

1. Sohraby, K., Minoli, D., Occhiogrosso, B., et al.: A review of wireless and satellite based M2M/IoT services in support of smart grids. Mob. Netw. Appl. **4**, 1–15 (2017)
2. Wietfeld, C., Georg, H., Groening, S., et al.: Wireless M2M communication networks for smart grid applications. In: Wireless Conference 2011–Sustainable Wireless Technologies, VDE, pp. 1–7 (2011)
3. Minoli, D.: Innovations in Satellite Communication and Satellite Technology: The Industry Implications of DVB-S2X, High Throughput Satellites, Ultra HD, M2M, and IP. Wiley (2015)
4. Kumar, Y.A., Sachin, T.: A tree based multicast routing protocol using reliable neighbor node for wireless mobile ad-hoc networks. In: Proceedings of the 4th International Conference on Frontiers in Intelligent Computing: Theory and Applications (FICTA) 2015. Springer, India (2016)
5. Helfert, M., Krempels, K.H., Klein, C., et al.: Smart cities, green technologies, and intelligent transport systems. Commun. Comput. Inf. Sci. (2015)
6. Lakhtaria, K.I.: Next Generation Wireless Network Security and Privacy (2015)
7. Beidou Satellite Navigation System. http://www.beidou.gov.cn/
8. Vassaki, S., et al.: Satellite-based sensor networks: M2M sensor communications and connectivity analysis. In: International Conference on Telecommunications and Multimedia, pp. 132–137, IEEE (2014)

Network Packet Capturing System for Communication Simulation in Smart Grid

Qianyu Ji[1], Yibo Yang[1], JunJing Zhang[1], Yunsong Yan[2], Zhijian Zhang[1], and Jian Wang[1(✉)]

[1] School of Electronic Science and Engineering, Nanjing University, 163 Xianlin Road, Nanjing 210023, Jiangsu, China
qinsheng_ji@foxmail.com, {1712116640,435080913}@qq.com, {njuzzj,wangjnju}@nju.edu.cn

[2] State Key Laboratory of Smart Grid Protection and Control, NARI Group Corporation, 19 Chengxin Road, Nanjing 211000, Jiangsu, China
448666225@qq.com

Abstract. The next-generation electric power systems (smart grid) are studied to solve the energy crisis. Communication system simulation for the smart grid is of great significance. In this paper, we introduce a simulation system of a real power system. Meanwhile, we established a packet capture system to monitor the situation of the network and tested its performance under different bandwidths. We discussed the reasons that might influence the system's performance and put forward some proposals for future work.

Keywords: Smart grid · Power system simulation · Network packet capture

1 Introduction

The current electric power systems have been serving us for decades. Next-generation power systems are endowed with automated and intelligent management, which is a critical component that determines the effectiveness and efficiency of the power systems. The concept *Smart Grid* was proposed several years ago, which has a variety of advantages over the current systems in terms of digitalization, flexibility, sustainability, and intelligence [1]. A fast, reliable, and secure communication network plays a vital role in the power system management.

With the development of abovementioned technologies, complex simulation tools become more and more significant. Thus, it is extremely meaningful to study power system communication simulations. As one of the enabling technologies, network packet capturing is of great practical significance to capture and analyze network data. This not only helps regulate and optimize the network, but also discovers potential network security issues and does preventive work on these issues [2]. However, most of the existing simulation systems neglect the

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 891–898, 2019.
https://doi.org/10.1007/978-981-13-6264-4_105

importance of an efficient capture system. Existed packet capture tools sometimes can be tedious and not portable enough. Therefore, it makes sense to customize a packet capture tool for simulation systems.

The rest of this paper is organized as follows. Section 2 demonstrates related work in smart grid and power system simulation. Section 3 presents current simulation systems including SCADA system and software-defined networking for smart grid communications. Section 4 introduces how the network capture system works and Sect. 5 demonstrates result of the experiments. Section 6 draws some conclusions and makes a summary of this paper.

2 Relative Work

The communication infrastructure in the smart grid must support the expected smart grid functionalities and meet the performance requirements. As the infrastructure connects an enormous number of electric devices and manages the complicated device communications, it is constructed in a hierarchical architecture with interconnected individual subnetworks and each taking responsibility of separate geographical regions [3]. An illustrative example of this architecture is shown in Fig. 1 [4].

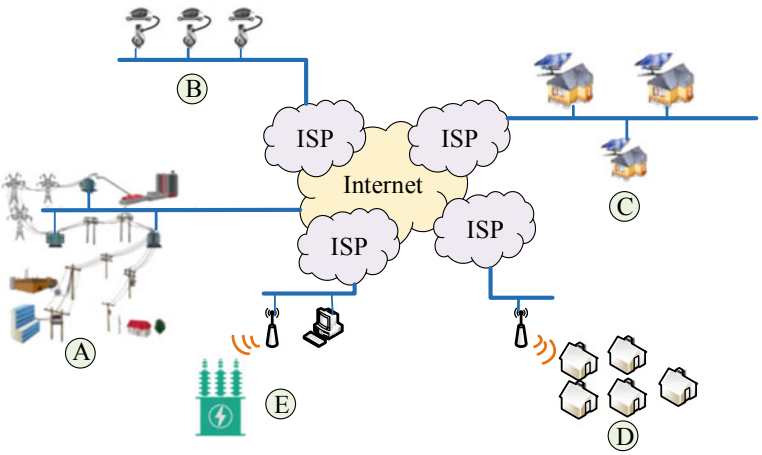


Fig. 1. An example of communication architecture in the smart grid. A is a segment of power transmission lines, B is a PEV charging station, and C is a residential subdivision installed with solar panels. D is a residential complex with AMI, and E is a power substation. The internet and ISPs serve as the backbone in connecting the distributed subnetworks.

Smart grid is a space-time multi-dimensional heterogeneous system in which power system and information communication system are deeply integrated [5]. The hybrid simulation of these points has become a research hotspot in recent

years. In order to combine the two types of simulation methods to study the relevant characteristics of the smart grid, researchers proposed the following three types of solutions: simultaneous simulation solutions, non-real-time hybrid simulation solutions, and implementation of hybrid simulation solutions.

The main idea of the simultaneous hybrid simulation scheme is to establish a complex power and information communication compound system model in a single simulation tool. Commonly used programs include integration of MATLAB program modules into OMNeT++ and other methods to perform research on issues related to demand response [6].

Another solution for the simulation of power communication complex system is to use hybrid simulation. The modeling work of the two systems still uses their own professional simulation software. To make the two softwares run at the same time domain, time synchronization method is utilized. Common tools include PSCAD/EMTDC & JAVA proposed by Mesut Baran [7], EPOCHS and VPNET proposed by Kenneth Hopkinson [8,9], PowerWorld Server & RINSE and TASSCS brought forward by C.M Davis [10], adevs & NS2 raised by James Nutaro [11], MATLAB/Simulink & OPNET used by Zhu et al. [12], and so on.

None of the above solutions take a real-time hybrid simulation into consideration. In 2014, Mikel Armendariz from KTH Royal Institute of Technology and the research group lead by Prof. Saifur Rahman from Virginia Tech separately proposed real-time hybrid simulation systems. They use a combination of RT-LAB and OPAL-RT, OPNET, and Java Eclipse to conduct their work [13].

3 Power System Communication Network

Future power systems are a complexity with numerous intelligent devices, exchanging huge amounts of data and real-time critical information. As information and communication networks are highly interconnected with the power grid, power system communication (PSC) has become a critical issue in the smart grid. Cyber security of a supervisory control and data acquisition (SCADA) system is researched to solve such issues.

In a SCADA system, a real power system communication network contains three main parts: the control center, power flow computing center, and the capture and analysis system. The power flow computing center includes the grid model and arithmetic program. Also, referring to the physical parameters used in the control system of the real power grid, it contains parameters like voltage level, phase angle, active power, and reactive power. Grid model contains power generation nodes and consuming nodes. The control center forwards data and commands. In each node, there are sensors and actuators to collect and send data. The web display subsystem is used to display the status and control information of each node in the simulation grid, interacts with the control center, and obtains the grid status at the current moment from the control center [14] (Fig. 2).

In the data acquisition plane, packets are captured in on the bottom layer, and on the upper layer, the system processes the data according to the different

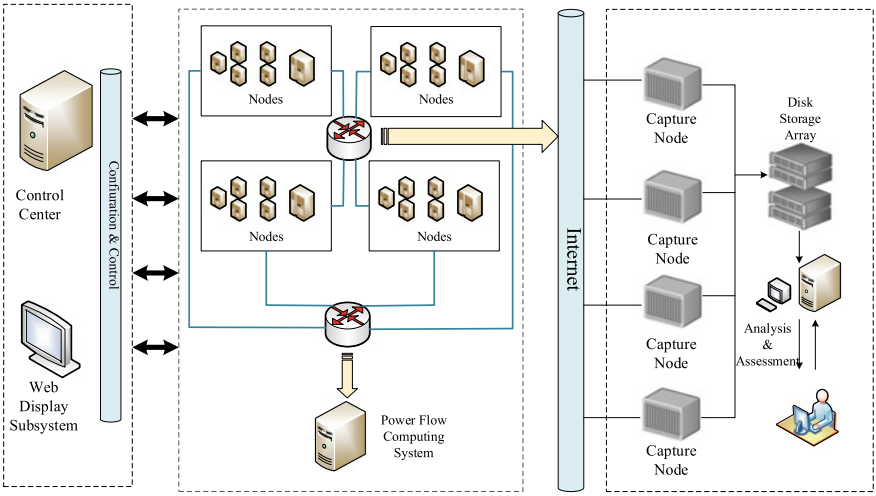


Fig. 2. An SCADA system. It contains three parts, the Control Center, the Power Flow Computing Center, and the Data Acquisition Plane

requirements of real-time display or data analysis (historical data). Mass data capture supports historical performance analysis. Comprehensive performance and alarm presentation are the overall external output of the entire assessment and test system.

4 Network Packet Capture

In the Linux kernel network protocol stack, when a packet is sent, the network packet arrives at the network device from the userspace via the network protocol stack, which is a process from top to bottom. Meanwhile, when a packet is received, it arrives at the network device and propagates upward through the network protocol stack and finally reaches the user space. The way that the capture system works is shown in Fig. 3.

When an external network packet enters the network adapter, the network adapter will write the arriving packet to the specified memory address through the direct memory access (DMA) method. This memory address is distributed and initialized by the network interface card driver, who is going to inform the CPU that there is data approaching through hardware interrupt request (IRQ). CPU invokes the corresponding function in the network card driver by querying the interrupt table. NIC driver will disable the IRQ, indicating the driver is aware of the packet to avoid interrupting the CPU for multiple times, in which way the efficiency is improved. Because IRQ cannot be interrupted, soft interrupt request is promoted, in order to prevent the CPU interrupt time from being too long to respond to other hardware devices.

In the kernel, there is a process ksoftirqd that is responsible for handling soft interrupt requests. When it receives a soft interrupt, it calls the corresponding

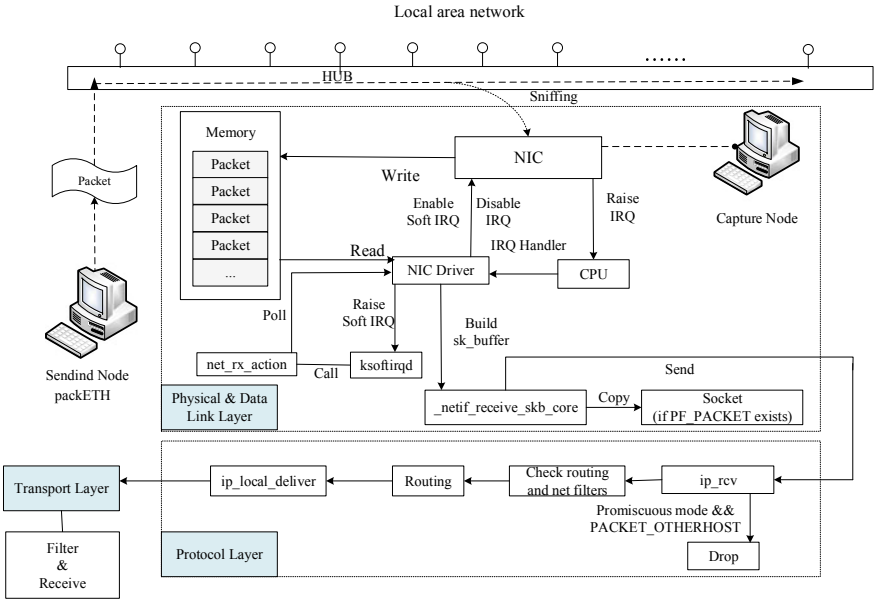


Fig. 3. The network capture system.

functions to process the request. In this case, the function poll in the network interface card driver is called to read the received data in the storage area and allocate the sk_buff data structure. Set the appropriate sk_buff parameter and covert to the skb format that can be recognized by the kernel network module. Call the relevant NAPI function to send the configured sk_buff structure upwards to the Linux kernel's network protocol stack. At the same time, detect whether there is a PF_PACKET type socket, if there is, copy the data to the socket. When all packets in the memory are processed, the network interface card driver starts the IRQ in the network card and wait for the arrival of the next packet.

The entrance function on the protocol layer is ip_rcv(), who processes packets first and dump packets that are useless. Packets left behind are sent to the next function ip_router_input(). In this function, if the destination IP address of the packet is not the local IP address and the IP forward function is not enabled, the packet is dumped. Otherwise. it enters the forwarding process, who handles TTL and Netfilter hook. After that, the packet is sent to the data link layer. If the packet's destination IP address is consistent with the local IP address, system calls ip_local_deliver() and send the packet to the transport layer, where decides the protocol of packets, does filter work, and resolve the packets.

5 Results and Discussions

5.1 Test Environment

All packets are transmitted on the bus in shared mode. In order to better test the capture system, the test environment is simplified to two nodes. On the sending node, use packETH to send packets. PackETH is a software that could control specific parameters of packets such as the length, source, and destination IP address. Meanwhile, packETH is also able to control the bandwidth and total numbers of packets to send. On the capture node, the network interface card is configured to the promiscuous mode. Libpcap library is used to find and open the network interfaces. Berkley Packet Filter (BPF) offers high-speed kernel filter for the system.

5.2 Results

Packets are saved as a pcap file for subsequent analysis. This process starts at the data link layer and includes data at every layer up. When sending node is turned on, capture node obtains packets that are sent from it. These packets have the same format. When the capture node visits the internet, Internet packets go through its network card and get caught. These packets have unequal length and contents. When the BPF filter is configured, packets that meet the requirements are received while others are dumped. When the filter is configured as "des host 192.168.1.102", which is the address of the capture node, packets that are sent to it are captured. If the filter is "proto tcp", TCP packets are received and ICMP, ARP, UDP packets are thrown away. Count the 20,000 packets that are captured from the internet and classify them according to the protocols. 86.6% of them are TCP packets, 3.4% are ICMP packets, 0.1% are ARP packets and 8.7% of them are UDP packets.

To discuss the performance of the system under different bandwidths guarantees that the system works well in different environments, as well as provides directions for improvements. When testing, packets are restricted to a length of 60 bytes. Change the bandwidths of sending packets to test the capture node's CPU usage and packet loss rate. In all conditions, 500,000 packets are sent.

As is demonstrated in Table 1, the system works well under a low bandwidth. When the bandwidth rises, CPU usage shows a trend of rising first and then decreasing. The packet loss rate of the system in the low-speed network is kept at zero. However, when the bandwidth is over 70 Mbps, its performance begins to be unsatisfactory. We assume that there are two reasons that caused such a problem. When the CPU is interrupted, some packets are dropped at such time. What is more, the copy of the packets from the kernel space to the user program occupies the computer resources and take time, which will increase the packet loss rate.

Table 1. System performance under different bandwidths

Bandwidth (Mbps)	CPU usage (%)	Packet loss rate (%)
10	8.7	0
20	18.3	0
30	25.9	0
40	36.3	0
50	43.9	0
60	53.0	0
70	52.8	8.27
80	49.7	10.03
90	49.8	11.72
100	48.2	14.55

6 Conclusions

The next-generation electric power system is expected to alleviate the energy shortage problem by exploiting renewable energy resources. In current power system simulation methods, the application of packet capture technology is rarely considered. Wireshark and other tools are used for system monitoring. These methods cannot perform function customization according to different simulation systems and cannot achieve real-time display effects. This article completes the preliminary work of establishing a capture mechanism for the simulation system. It focuses on the application space of data packet capture technology in power system simulation, and carries out system simulation, tests the capture technology, and implements packet capture, filter setup, packet analysis, and performance testing.

Acknowledgement. This work was supported by State Key Laboratory of Smart Grid Protection and Control of NARI Group Corporation.

References

1. Li, F., Qiao, W., Sun, H., Wan, H., Wang, J., Xia, Y., et al.: Smart transmission grid: vision and framework. IEEE Trans. Smart Grid **1**(2), 168–177 (2010)
2. Baronti, F., Vazquez, S., Chow, M.Y.: Modeling, control, and integration of energy storage systems in e-transportation and smart grid. IEEE Trans. Ind. Electron. **65**(8), 6548–6551 (2018)
3. Gungor, V.C., Lambert, F.C.: A survey on communication networks for electric system automation. Comput. Netw. **50**(7), 877–897 (2006)
4. Wang, W., Xu, Y., Khanna, M.: A survey on the communication architectures in smart grid. Comput. Netw. **55**(15), 3604–3629 (2011)

5. Mets, K., Ojea, J.A., Develder, C.: Combining power and communication network simulation for cost-effective smart grid analysis. IEEE Commun. Surv. Tutor. **16**(3), 1771–1796 (2014)
6. Mets, K., Verschueren, T., Develder, C., Vandoorn, T.L., Vandevelde, L.: Integrated simulation of power and communication networks for smart grid applications. In: 2011 IEEE 16th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), pp. 61–65
7. Baran, M., Sreenath, R., Mahajan, N.R., Baran, M., Sreenath, R., Mahajan, N.R.: Extending EMTDC/PSCAD for simulating agent-based distributed applications. IEEE Power Eng. Rev. **22**(12), 52–54 (2002)
8. Hopkinson, K., Wang, X., Giovanini, R., Thorp, J., Birman, K., Coury, D.: EPOCHS: a platform for agent-based electric power and communication simulation built from commercial off-the-shelf components. IEEE Trans. Power Syst. **21**(2), 548–558 (2006)
9. Li, W., Monti, A., Luo, M., Dougal, R.A.: Vpnet: a co-simulation framework for analyzing communication channel effects on power systems. In: Electric Ship Technologies Symposium, pp. 143–149
10. Bian, D., Kuzlu, M., Pipattanasomporn, M., Rahman, S., Wu, Y.: Real-time co-simulation platform using OPAL-RT and OPNET for analyzing smart grid performance, pp. 1–5
11. Mallouhi, M., Al-Nashif, Y., Cox, D., Chadaga, T., Hariri, S.: A testbed for analyzing security of SCADA control systems (TASSCS). In: ISGT 2011, pp. 1–7
12. Zhu, K., Chenine, M., Nordstrom, L.: Ict architecture impact on wide area monitoring and control systems' reliability. IEEE Trans. Power Deliv. **26**(4), 2801–2808 (2011)
13. Armendariz, M., Chenine, M., Nordstrm, L., Al-Hammouri, A.: A co-simulation platform for medium/low voltage monitoring and control applications. In: ISGT 2014, pp. 1–5
14. Dao, P.B., Staszewski, W.J., Barszcz, T., Uhl, T.: Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data. Renew. Energy (2017)

Diversity Gain of Millimeter-Wave Massive MIMO Systems with Distributed Antenna Arrays

Dian-Wu Yue[1(✉)], Shuai Xu[1], and Ha H. Nguyen[2]

[1] College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China
dwyue@dlmu.edu.cn

[2] Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon S7N 5A9, Canada

Abstract. This paper is concerned with the diversity performance analysis for millimeter-wave (mmWave) massive MIMO systems. For a single-user mmWave system employing distributed antenna-subarray architecture in which the transmitter and receiver consist of K_t and K_r subarrays, respectively, a diversity gain theorem is established when the numbers of antennas at subarrays go to infinity. Specifically, assuming that all subchannels have the same number of propagation paths L, the theorem states that by employing such a distributed antenna-subarray architecture, a diversity gain of $K_r K_t L - N_s + 1$ can be achieved, where N_s is the number of data streams. This result implies that compared to the colocated antenna architecture, using the distributed antenna-subarray architecture can scale up the diversity gain or multiplexing gain proportionally to $K_r K_t$.

Keywords: Millimeter wave · Massive MIMO · Diversity gain · Multiplexing gain

1 Introduction

Millimeter-wave (mmWave) communication has recently gained considerable attention as a candidate technology for 5G mobile communication systems and beyond [1]. The main reason for this is the availability of vast spectrum in the mmWave band that is very attractive for high data rate communications. However, compared to communication systems operating at lower microwave frequencies, propagation loss in mmWave frequencies is much higher, in order of magnitude. Fortunately, given the much smaller carrier wavelengths, mmWave communication systems can make use of compact massive antenna arrays to compensate for the increased propagation loss. Nevertheless, the large-scale antenna arrays together with high cost and large power consumption of the mixed analog/digital signal components make it difficult to equip a separate

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 899–904, 2019.
https://doi.org/10.1007/978-981-13-6264-4_106

radio-frequency (RF) chain for each antenna element and perform all the signal processing in the baseband. Therefore, research on *hybrid* analog–digital processing of precoder and combiner for mmWave communication systems have attracted very strong interests from both academia and industry [2–4].

Due to the fact that the antenna arrays in the abovementioned RF architectures are colocated and mmWave signal propagation has an important feature of multipath sparsity in both the temporal and spatial domains [5], it is expected that the potentially available diversity and multiplexing gains are not large for the colocated antenna deployment. In order to enlarge diversity/multiplexing gains in mmWave massive MIMO communication systems, this paper considers a more general antenna array architecture, called *distributed antenna-subarray architecture*, which includes colocated array architecture as special cases. For a mmWave massive MIMO system, how to quantify the diversity/multiplexing performance is a fundamental and open research problem [6]. In particular, to the best of our knowledge, until now there is no unified diversity gain analysis for mmWave massive MIMO systems that are applicable to both colocated and distributed antenna array architectures. To fill this gap, this paper investigates the diversity performance of mmWave massive MIMO systems with the proposed distributed subarray architecture. The focus is on the asymptotical diversity gain analysis in order to find out the potential diversity advantage provided by multiple distributed antenna arrays.

2 System Model

Consider a single-user mmWave massive MIMO system as shown in Fig. 1. The transmitter is equipped with a distributed antenna array to send N_s data streams to a receiver, which is also equipped with a distributed antenna array. Here, a distributed antenna array means an array consisting of several remote antenna units (RAUs) (i.e., antenna subarrays) that are distributively located. Specifically, the antenna array at the transmitter consists of K_t RAUs, each of which has N_t antennas and is connected to a baseband processing unit (BPU) by fiber. Likewise, the distributed antenna array at the receiver consists of K_r RAUs, each having N_r antennas and also being connected to a BPU by fibers. The

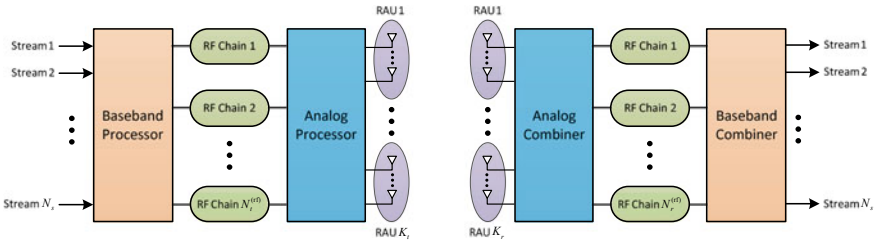


Fig. 1. Block diagram of a mmWave massive MIMO system with distributed antenna arrays.

transmitter accepts as its input N_s data streams and is equipped with $N_t^{(\mathrm{rf})}$ RF chains, where $N_s \leq N_t^{(\mathrm{rf})} \leq N_t K_t$. Given $N_t^{(\mathrm{rf})}$ transmit RF chains, the transmitter can apply a low-dimensional $N_t^{(\mathrm{rf})} \times N_s$ baseband precoder, $\mathbf{W}_t$, followed by a high-dimensional $K_t N_t \times N_t^{(\mathrm{rf})}$ RF precoder, $\mathbf{F}_t$. Note that, the amplitude and phase modifications are feasible for the baseband precoder $\mathbf{W}_t$, while only phase changes can be made by the RF precoder $\mathbf{F}_t$ through the use of variable phase shifters and combiners. The transmitted signal vector can be written as

$$\mathbf{x} = \mathbf{F}_t \mathbf{W}_t \mathbf{s}, \tag{1}$$

where $\mathbf{s}$ is the $N_s \times 1$ symbol vector such that $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \frac{P}{N_s}\mathbf{I}_{N_s}$. Thus, P represents the average total input power. Considering a narrowband block fading channel, the $K_r N_r \times 1$ received signal vector is

$$\mathbf{y} = \mathbf{H}\mathbf{F}_t \mathbf{W}_t \mathbf{s} + \mathbf{n} \tag{2}$$

where $\mathbf{H}$ is $K_r N_r \times K_t N_t$ channel matrix and $\mathbf{n}$ is a $K_r N_r \times 1$ vector consisting of i.i.d. $CN(0,1)$ noise samples. Throughout this paper, $\mathbf{H}$ is assumed known to both the transmitter and receiver. Given that $N_r^{(\mathrm{rf})}$ RF chains (where $N_s \leq N_r^{(\mathrm{rf})} \leq N_r K_r$) are used at the receiver to detect the N_s data streams, the processed signal is given by

$$\mathbf{z} = \mathbf{W}_r^H \mathbf{F}_r^H \mathbf{H}\mathbf{F}_t \mathbf{W}_t \mathbf{s} + \mathbf{W}_r^H \mathbf{F}_r^H \mathbf{n} \tag{3}$$

where $\mathbf{F}_r$ is the $K_r N_r \times N_r^{(\mathrm{rf})}$ RF combining matrix, and $\mathbf{W}_r$ is the $N_r^{(\mathrm{rf})} \times N_s$ baseband combining matrix. Furthermore, according to the architecture of RAUs at the transmitting and receiving ends, $\mathbf{H}$ can be written as

$$\mathbf{H} = \begin{bmatrix} \sqrt{g_{11}}\mathbf{H}_{11} & \cdots & \sqrt{g_{1K_t}}\mathbf{H}_{1K_t} \\ \vdots & \ddots & \vdots \\ \sqrt{g_{K_r1}}\mathbf{H}_{K_r1} & \cdots & \sqrt{g_{K_rK_t}}\mathbf{H}_{K_rK_t} \end{bmatrix}. \tag{4}$$

In the above expression, g_{ij} represents the large-scale fading effect between the ith RAU at the receiver and the jth RAU at the transmitter, which is assumed to be constant over many coherence-time intervals. The normalized subchannel matrix $\mathbf{H}_{ij}$ represents the MIMO channel between the jth RAU at the transmitter and the ith RAU at the receiver.

A clustered channel model based on the extended Saleh-Valenzuela model is often used in mmWave channel modeling and standardization [3] and it is also adopted in this paper. For simplicity of exposition, each scattering cluster is assumed to contribute a single propagation path. Using this model, the subchannel matrix $\mathbf{H}_{ij}$ is given by

$$\mathbf{H}_{ij} = \sqrt{\frac{N_t N_r}{L_{ij}}} \sum_{l=1}^{L_{ij}} \alpha_{ij}^l \mathbf{a}_r(\phi_{ij}^{rl}, \theta_{ij}^{rl}) \mathbf{a}_t^H(\phi_{ij}^{tl}, \theta_{ij}^{tl}), \tag{5}$$

where L_{ij} is the number of propagation paths, α_{ij}^{l} is the complex gain of the lth ray, and ϕ_{ij}^{rl} (θ_{ij}^{rl}) and ϕ_{ij}^{tl} (θ_{ij}^{tl}) are its random azimuth (elevation) angles of arrival and departure, respectively. Without loss of generality, the complex gains α_{ij}^{l} are assumed to be $CN(0,1)$. The vectors $\mathbf{a}_r(\phi_{ij}^{rl},\theta_{ij}^{rl})$ and $\mathbf{a}_t(\phi_{ij}^{tl},\theta_{ij}^{tl})$ are the normalized receive/transmit array response vectors at the corresponding angles of arrival/departure. For an N-element uniform linear array (ULA) , the array response vector is

$$\mathbf{a}^{\mathrm{ULA}}(\phi)=\frac{1}{\sqrt{N}}\left[1,\mathrm{e}^{j2\pi\frac{d}{\lambda}\sin(\phi)},\ldots,\mathrm{e}^{j2\pi(N-1)\frac{d}{\lambda}\sin(\phi)}\right]^{T} \tag{6}$$

where λ is the wavelength of the carrier and d is the inter-element spacing. It is pointed out that the angle θ is not included in the argument of $\mathbf{a}^{\mathrm{ULA}}$ since the response for an ULA is independent of the elevation angle. In contrast, for a uniform planar array (UPA), which is composed of N_h and N_v antenna elements in the horizontal and vertical directions, respectively, the array response vector is represented by

$$\mathbf{a}^{\mathrm{UPA}}(\phi,\theta)=\mathbf{a}_h^{\mathrm{ULA}}(\phi)\otimes\mathbf{a}_v^{\mathrm{ULA}}(\theta), \tag{7}$$

where

$$\mathbf{a}_h^{\mathrm{ULA}}(\phi)=\frac{1}{\sqrt{N_h}}\left[1,\mathrm{e}^{j2\pi\frac{d_h}{\lambda}\sin(\phi)},\ldots,\mathrm{e}^{j2\pi(N_h-1)\frac{d_h}{\lambda}\sin(\phi)}\right]^{T} \tag{8}$$

and

$$\mathbf{a}_v^{\mathrm{ULA}}(\theta)=\frac{1}{\sqrt{N_v}}\left[1,\mathrm{e}^{j2\pi\frac{d_v}{\lambda}\sin(\theta)},\ldots,\mathrm{e}^{j2\pi(N^v-1)\frac{d_v}{\lambda}\sin(\theta)}\right]^{T}. \tag{9}$$

Throughout this paper, we suppose that the antenna configurations at all RAUs are either ULA or UPA.

3 Diversity Gain Analysis

From the structure and definition of the channel matrix $\mathbf{H}$ in Sect. 2, there is a total of $L_s=\sum_{i=1}^{K_r}\sum_{j=1}^{K_t}L_{ij}$ propagation paths. Naturally, $\mathbf{H}$ can be decomposed into a sum of L_s rank-one matrices, each corresponding to one propagation path. Specifically, $\mathbf{H}$ can be rewritten as

$$\mathbf{H}=\sum_{i=1}^{K_r}\sum_{j=1}^{K_t}\sum_{l=1}^{L_{ij}}\tilde{\alpha}_{ij}^{l}\tilde{\mathbf{a}}_r(\phi_{ij}^{rl},\theta_{ij}^{rl})\tilde{\mathbf{a}}_t^{H}(\phi_{ij}^{tl},\theta_{ij}^{tl}), \tag{10}$$

where

$$\tilde{\alpha}_{ij}^{l}=\sqrt{g_{ij}\frac{N_tN_r}{L_{ij}}}\alpha_{ij}^{l}, \tag{11}$$

$\tilde{\mathbf{a}}_r(\phi_{ij}^{rl},\theta_{ij}^{rl})$ is a $K_rN_r\times 1$ vector whose bth entry is defined as

$$[\tilde{\mathbf{a}}_r(\phi_{ij}^{rl},\theta_{ij}^{rl})]_b=\begin{cases}[\mathbf{a}_r(\phi_{ij}^{rl},\theta_{ij}^{rl})]_{b-(i-1)N_r}, & b\in Q_i^r\\ 0, & b\notin Q_i^r\end{cases} \tag{12}$$

where $Q_i^r = ((i-1)N_r, iN_r]$. And $\tilde{\mathbf{a}}_t(\phi_{ij}^{tl}, \theta_{ij}^{tl})$ is a $K_tN_t \times 1$ vector whose bth entry is defined as

$$[\tilde{\mathbf{a}}_t(\phi_{ij}^{tl}, \theta_{ij}^{tl})]_b = \begin{cases} [\mathbf{a}_t(\phi_{ij}^{tl}, \theta_{ij}^{tl})]_{b-(j-1)N_t}, & b \in Q_j^t \\ 0, & b \notin Q_j^t. \end{cases} \tag{13}$$

where $Q_j^t = ((j-1)N_t, jN_t]$.

Lemma 1 *All L_s vectors $\{\tilde{\mathbf{a}}_r(\phi_{ij}^{rl}, \theta_{ij}^{rl})\}$ are orthogonal to each other when $N_r \to \infty$. Likewise, all L_s vectors $\{\tilde{\mathbf{a}}_t(\phi_{ij}^{tl}, \theta_{ij}^{tl})\}$ are orthogonal to each other when $N_t \to \infty$.*

Based on Lemma 1 and the knowledge of order statistics, we can derive the following result.

Theorem 1 *Let $N_s \leq L_s$. Then the distributed massive MIMO system with very large N_r and N_t can achieve a diversity gain of*

$$G_d = L_s - N_s + 1. \tag{14}$$

Remark 1: When N_t and N_r are large enough, (10) indicates that the system multiplexing gain is at most equal to L_s. This is reasonable since there exist only L_s effective singular values in the channel matrix $\mathbf{H}$. Theorem 1 provides a simple diversity-multiplexing trade-off of a mmWave massive MIMO system: adding one data stream to the system decreases the diversity gain by one, whereas removing one data stream increases the diversity gain by one. Such a trade-off is useful in designing a system to meet the requirements on both data rate and error performance.

Corollary 1 *Assume that $L_{ij} = L$ for any i and j. Let $N_s \leq K_rK_tL$. When both N_t and N_r are very large, the distributed massive MIMO system can achieve a diversity gain*

$$G_d = K_rK_tL - N_s + 1. \tag{15}$$

In particular, when $K_r = K_t = 1$, the massive MIMO system with colocated antennas arrays can achieve a diversity gain

$$G_d = L - N_s + 1. \tag{16}$$

Remark 2: Corollary 1 implies that for a mmWave colocated massive MIMO system, its diversity gain and multiplexing gain are limited and at most equal to the number of paths L. However, these gains can be increased by employing the distributed antenna architecture and can be scaled up proportionally to K_rK_t.

4 Conclusion

This paper has provided asymptotical diversity analysis for mmWave massive MIMO systems with colocated and distributed antenna architectures when the

number of antennas at each subarray goes to infinity. Theoretical analysis shows that with a colocated massive antenna array, scaling up the number of antennas of the array can increase the coding gain (array gain), but not the diversity gain. However, if the array is built from distributed subarrays (RAUs), each having a very large number of antennas, then increasing the number of RAUs does increase the diversity gain and/or multiplexing gain. As such, the analysis leads to a novel approach to improve the diversity and multiplexing gains of mmWave massive MIMO systems. It is acknowledged that the asymptotical diversity analysis obtained in this paper is under the idealistic assumption of having perfect CSI. Performing the diversity analysis for mmWave massive MIMO systems under imperfect CSI is important and deserves further research.

Acknowledgements. This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 3132016347 and The Natural Science Foundation of Liaoning Province under Grant 201602086.

References

1. Roh, W., et al.: Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results. IEEE Commun. Mag. **52**(2), 106–113 (2014)
2. Sohrabi, F., Yu, W.: Hybrid digital and analog beamforming design for large-scale antenna arrays. IEEE J. Sel. Top. Signal Process. **10**(3), 501–513 (2016)
3. Ayach, O.E., et al.: The capacity optimality of beam steering in large millimeter wave MIMO systems. In: IEEE Workshop on Signal Processing Advances in Wireless Communications, pp. 100–104 (2012)
4. Molisch, A.F., et al.: Hybrid beamforming for massive MIMO-a survey. IEEE Commun. Mag. **55**(9), 134–141 (2017)
5. Rappaport, T.S., et al.: Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications. IEEE Trans. Antennas Propag. **61**(4), 1850–1859 (2013)
6. Zheng, L., Tse, D.N.C.: Diversity and multiplexing: a fundamental tradeoff in multiple antenna channels. IEEE Trans. Inform. Theory **49**(5), 1073–1096 (2003)

Distributed Power Control for Energy Conservation in Hybrid Cellular Network with Device-to-Device Communication

Yafei Yang[1,2], Yuhan Dai[1,2], Hui Li[1,2(✉)], and Ruiqing Zhang[1,2]

[1] College of Information Science and Technology, Hainan University, Haikou, China
yafei94@yeah.net, {dyh930910,hitlihuillll2, iezhangruiqing}@163.com
[2] Marine Communication and Network Engineering Technology Research Center of Hainan Province, Hainan University, Haikou, China

Abstract. The device-to-device (D2D) communication is the two terminals that can communicate directly and be controlled by the system. It can improve the efficiency of the cellular network communication greatly, but the interference caused by D2D communication has brought a series of problems to cellular networks' management. In order to limit the interference of cross-layer and improve the utilization rate of hybrid cellular networks, we design a new kind of distributed power control (DPC) scheme. On the premise of meeting the requirements of quality of service, we further increased a biasing scheme into DPC, to avoid invalid power consumption, so as to realize the purpose of energy saving. The feasibility of the scheme is proved by the comparison and analysis of the simulation results.

Keywords: D2D communication · Distributed power control · Energy saving · Biasing scheme

1 Introduction

In today's society, smart user devices are becoming more and more popular, and the demand for wireless Internet access and quality of service (QoS) is increasing. With the increasing demands of battery and energy efficiency, information and communication technology is facing with a series of problem including how to reduce gas emissions in the greenhouse. So, green communication [1–3] received extensive attention from industry and academia.

D2D communication [4, 5] is a kind of new communication mode, by reusing the existing cellular network resources, in which two D2D users no longer depend on the base station and could build direct data link. So that people can shorten the transmission delay and even reduce the burden of the base station. In view of the advantages and characteristics of D2D communication, it has been widely studied as a direction for green communication. But due to the reuse of spectrum, serious interference comes from the D2D users (D-users in D-layer) or cellular user (C-users in S-layer). The D2D communication cannot send and receive signals without any limitation and restriction.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 905–913, 2019.
https://doi.org/10.1007/978-981-13-6264-4_107

How to limit the disruption of across layers and reduce the user's transmission power is a thorny issue. Distributed power control can manage the interference of the channel and optimize the performance of the system, which has been proven to be an effective way to implement the QoS requirements of the user [7–9].

2 Traditional Distributed Power Control Scheme

This part is based on the minimum balance power control (MBPC). The distributed power control scheme, which is based on the minimum signal-to-interference-plus-noise ratio (SINR), is changed with the user requirement and its delicate design is based on the following formula:

$$P_i[m+1] = P_i[m]\frac{\Gamma_i^{tar}}{\Gamma[m]} = \Gamma_i^{tar}\left(\sum_{j\neq k} P_j[m]G_{jk} + \delta_i\right)/G_{jk} \tag{1}$$

where P_i is the transmit power of the i^{th}, P_j is the transmit power of the j^{th}, Γ_i is the i^{th} a link of SINR, and the symbol in brackets indicates the number of iterations. Γ_i^{tar} is the required SINR target, G_{jk} is the between path gain j^{th} transmitter and the k^{th} receiver (j, k = 1, 2, 3…), and δ_i is the power spectral density of noise. We define the total ratio of interference plus noise to the path gain in the desired link as the effective interference $\gamma_i[m]$, which is $\gamma_{\text{i}}[m]=\left(\sum_{\text{j}\neq k} P_j[m]G_{jk} + \delta_i\right)/G_{jk}$, and the process of power updating can be simplified

$$I_{\text{i}}^{1}(P) = P_i[m+1] = \Gamma_i^{tar}\gamma_{\text{i}}[m] \tag{2}$$

We assume that other conflicting users transmit the same power before the next iteration as shown in Formula (1). One can see that once the actual SINR levels are not consistent with the target, the transmitting power will be updated to maintain the target signal-to-noise ratio. In addition, the power update process in Formula (1) can be briefly described as

$$\min_{P_i}\sum_i P_i, s.t.\Gamma_i \geq \Gamma_i^{tar}, \forall i \tag{3}$$

The optimization problem described in Formula (3) minimizes the total power consumption while achieving all the SINR requirements. In order to satisfy the actual requirement, a maximum power constraint is added to the Formula (2), and the process of power updating can be described as

$$I_{\text{i}}^{1}(P) = P_i[m+1] = \min\left\{\Gamma_i^{tar}\gamma_{\text{i}}[m], P_{MAX}\right\} \tag{4}$$

where the P_{MAX} is maximum transmit power. In [10], a group of DPC programs was analyzed, and the definition of the standard function includes (1) and (4) in two special circumstances.

This part is based on the distribute conflict power control (DCPC). Low SINR level cannot reach the QoS requirements, and the high level of SINR can lead to system capacity and infeasibility. All in all, the wrong match SINR targets and the global channel conditions will result in inadequate use of radio resources. Inspired by this, a new scheme was proposed in the literature [11], in which the update process of emission power was based on

$$I_{\mathrm{i}}^2(P) = P_i[m+1] = \theta/\gamma_i[m] = \left(\theta/\gamma_i^2[m]\right)\gamma_i[m] \tag{5}$$

where θ is defined as the signal interference factor, indicating the transmission request, say in other words. In the same $\gamma_i[m]$, the larger the value of θ, the better the conditions of the channel, and therefore can withstand greater power. In addition, we derive $\theta/\gamma_i^2[m] = \Gamma_i^{tar}$ from the derivation of Formula (5). We can see that the updated parameter of the transmitted power $\gamma_i[m]$ becomes the reciprocal. If $\gamma_i[m]$ increases, the user will reduce his SINR target and avoid the disproportionate power consumption; if not, the user will increase his SINR target to make full use of the channel conditions to improve wireless resource utilization.

Distributed power control schemes based on conflicting user diversity can avoid power divergence, and each user can set their own target SINR in a timely and reasonable manner, thus increasing the capacity of the system. But it cannot guarantee a minimum signal-to-noise ratio threshold, even if this threshold can be easily achieved. In contrast to Formula (4), the distributed power control scheme based on the conflicting user diversity can be summarized as

$$I_{\mathrm{i}}^2(P) = P_i[m+1] = \min\{\theta/\gamma_{\mathrm{i}}[m], P_{MAX}\} \tag{6}$$

3 System Model and Simulation Model

We simply consider a cellular solution for analysis as shown in Fig. 1. Assuming that the D-layer only multiplexes the uplink resources, and only the uplink is considered at the S-layer.

The base station has N D links in the center of the cell, and a single S-layer user is randomly and independently located in the cell. The minimum and maximum distances for each D-layer link are $d_{\min}^{link}$ and $\mathrm{d}_{\max}^{link}$. QoS is abstracted and measured by simplified SINR, and the SINR of the i^{th} user is defined as

$$\Gamma_i = \frac{G_{ii}P_i}{\eta_i + \sum_{j=0, j\neq k}^{N} G_{jk}P_j} \tag{7}$$

where P_i the transmit power of the i^{th} transmitter, the maximum transmit power is P_{MAX}, G_{ii} is the path gain of the i^{th} ideal link and G_{jk} is the gain of the path from j^{th} transmitter and the k^{th} receiver (j, k = 0, 1, 2, 3…), η_i the power spectral density, and the minimum SINR threshold is Γ_i^{thr}. The maximum SINR $\Gamma_{\max}$ is limited by the modulation and coding scheme. No additional data rates are added to the interference. The SINR of the i^{th} link satisfies $\Gamma_i^{thr} \le \Gamma_i \le \Gamma_{\max}$. Each link is looked on as a Gaussian channel with Shannon capacity. The capacity of the i^{th} link satisfies

$$C_i = W \log_2^{(1+\Gamma_i)}, (0 \le i \le N) \tag{8}$$

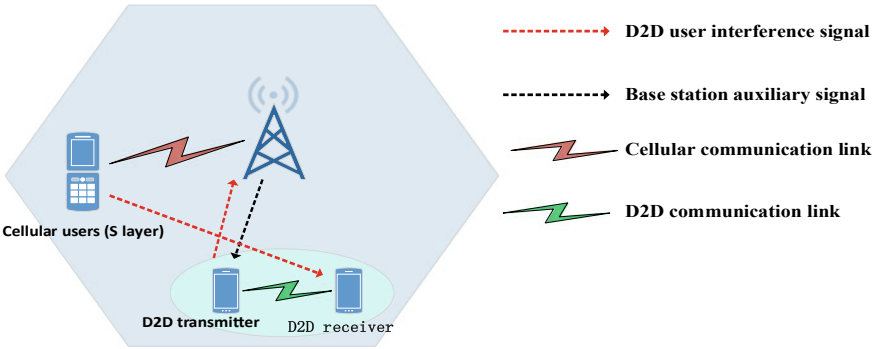


Fig. 1. System model scene diagram

We assume that both the base station and all users use omnidirectional antennas, and the antenna gain and noise figure for base stations and users are λ_{BS}, NF_{BS}, λ_{UE}, and NF_{UE}. It is assumed that all channel gains are deterministic (time averaged) and are not affected by channel fluctuations. If the receiver is the base station, the case of A in Formula (9) is adopted; and if the receiver is the user, the case of B in Formula (9) is adopted

$$\begin{cases} L_{ij} = 128.1 + 37.6 \lg(D_{ij}), i = 0 \quad \text{A} \\ L_{ij} = 148.2 + 40 \lg(D_{ij}), i \neq 0 \quad \text{B} \end{cases} \tag{9}$$

where D_{ij} is the distance between the j^{th} transmitter and the i^{th} receiver, in kilometers.

We develop our scheme on such specific topology to better illustrate their characteristics. Our scheme is not limited to the selected topology but is suitable for more general cases. The main simulation parameters are listed and set in Table 1.

Table 1. Basic parameters utilized in the simulation

Cell radius	R = 300 m
D2D link distance range	$d_{\min}^{link} = 28$ m, $d_{\max}^{link} = 45$ m
Maximum SINR level	$\Gamma_{\max} = 22$ dB
Antenna gain	$\lambda^{UE} = 0$ dBi, $\lambda^{BS} = 10$dBi
Maximum transmit power	$P_{MAX} = 150$ mW
Background noise	PSDbn = −174 dBm/Hz
Channel bandwidth	W = 180 kHz
Noise figure	$NF^{UE} = 10$ dB, $NF^{BS} = 4$ dB

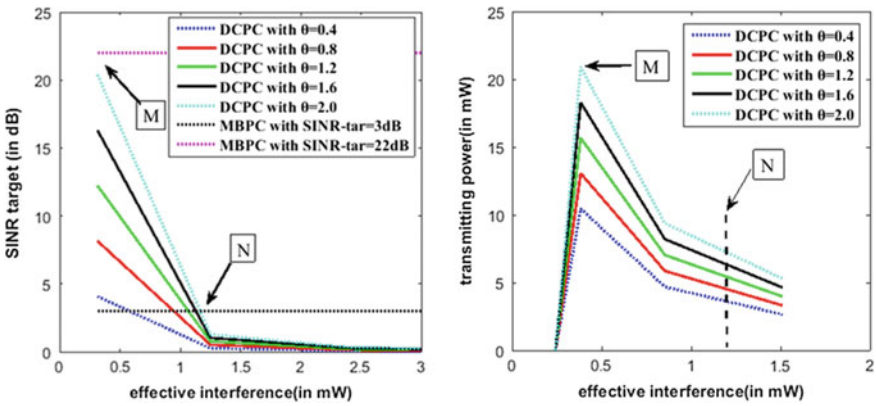
4 Distributed Power Control Scheme

This part we establish the constrained distribute conflict power control (CDCPC). Based on the features of MBPC and DCPC, we establish a distributed power control model with constrained DCPC,

$$I^3 = \min\{\max\{\min\{I_1^1(P), I_1^2(P)\}, I_2^1(P)\}, P_{MAX}\} \tag{10}$$

where $I_{1\,i}^1(P) = \Gamma_{MAX}\gamma_i[m], I_{2\,i}^1(P) = \Gamma_i^{thr}\gamma_i[m], I_1^2(P) = \theta/\gamma_i[m]$.

This scheme preserves the features of DCPC while striving to meet QoS requirements and to comply with constraints of the system architecture. In Fig. 2(a) and (b), point M and point N are defined as effective interference A and B, and their intersection is based on the DCPC scheme and two MBPC schemes. So, it can be easily deduced that $\gamma_i^M = \sqrt{\theta/\Gamma_{MAX}}$ and $\gamma_i^N = \sqrt{\theta/\Gamma_i^{thr}}$.



(a) The SINR target vs. effective interference

(b) Transmission power vs. effective interference

Fig. 2. SINR target and transmit power distribution of CDCPC scheme.

CDCPC divides the curve between γ_i and SINR into four stages. In the first stage $\left(\gamma_i \le \gamma_i^M\right)$, the channel conditions are very good, no additional power consumption, and the user can use the minimum signal-to-noise ratio based on the MBPC by the maximum signal-to-noise ratio $\Gamma_{MAX.}$ In the second stage $\left(\gamma_i^M < \gamma_i \le \gamma_i^N\right)$, there is certain interference, one can adapt its SINR level to γ_i in DCPC. In the third stage $\left(\gamma_i^N < \gamma_i \le \gamma_i^Q\right)$, where $\gamma_i^Q = \mathrm{P}_{MAX}/\Gamma_i^{thr}$, the channel condition is not good and the users only allow to enjoy the SINR threshold Γ_i^{thr} in MBPC. In the fourth stage $\left(\gamma_i > \gamma_i^Q\right)$, the channel condition is extremely bad and the users can only use the maximum power.

When the channel conditions are better, the D2D user tends to enjoy the high SINR and the power level corresponding to the first stage and the second stage; but when the D-layer transmitter is close to the base station or the system capacity increases, this will aggravate the degree of user interference in S-layer. Therefore, it will be difficult for the S-layer users to maintain the minimum threshold, in which the users enters the third stage or even the fourth stage, violates the original intention of the program design and leads to useless power consumption. In order to ensure the feasibility of S-layer users and to achieve the goal of energy-saving, S-layer users and D-layer users should be to coordinate with the system capacity.

In the CDCPC program, each D-level user sets its own SINR level separately, so that the channel conditions of the S-layer cannot be felt easily. Therefore, we introduce the cross-layer interference perception to the D-layer user and enlarge the definition of effective interference

$$\gamma_i^{\tau} = \gamma_i + P_i G_i^{BS}/G_{ii} = (1 + \Gamma_i \varpi_i)\gamma_i, \forall i \neq 0 \tag{11}$$

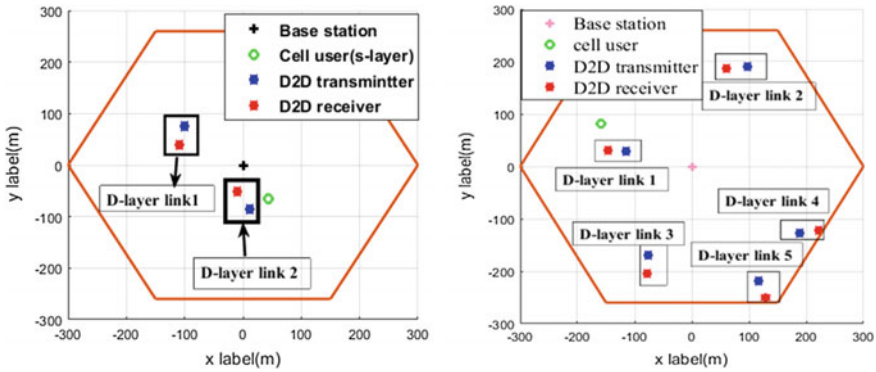
where G_i^{BS} is the path gain, in which path the i^{th} D-layer is linked to the base station; $P_i G_i^{BS}/G_{ii} = P_i \varpi_i$ indicates the degree of cross-layer interference and γ_i^{τ} is the combined interference. By comparing with Formula (5), a penalty factor of the signal interference $\theta_i^{bia} = \theta/(1 + \Gamma_i \varpi_i)$ is assigned to each D-layer link in the bias scheme. According to the severity of the cross-layer interference, the inefficient power consumption of the S-layer is reduced under adverse channel conditions, and it also weakens the total interference level. If $I_{2\mathrm{i}}^1(P) = \theta/\gamma_i^{\tau}$, the bias scheme can be defined as

$$I^4 = \min\{\max\{\min\{I_1^1(P), I_2^2(P)\}, I_2^1(P)\}, P_{MAX}\} \tag{12}$$

Since the D-layer's transmission request is limited not only by its own channel conditions but also by its cross-layer interference capability, the scheme is biased to the S-layer and indicates its priority in the hybrid system.

We set $\theta = 2,\ \Gamma_i^{thr} = 3dB, \forall i$, in both cases, the user conditions of the S-layer are in line with the third stage. The network topology is shown in Fig. 3.

In Fig. 3, the first topology simulates the power and signal-to-noise ratio of two D2D communication links. The second picture shows the power distribution in the case of increasing the number of users.



(a) Topology for two D-layer links

(b) Network topology of five D-layer links

Fig. 3. Specific network topology.

First of all, for comparative analysis, we simulated according to the network topology of the previous three programs. The simulation results are shown in Fig. 4.

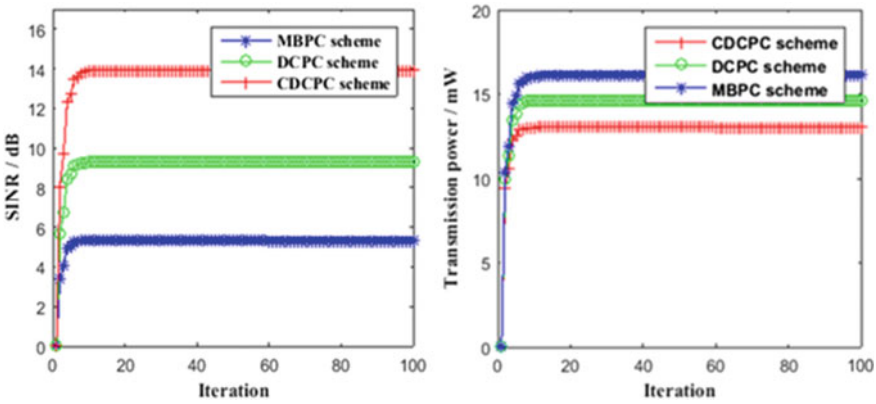


Fig. 4. Comparison of MBPC, DCPC, and CDCPC programs.

The traditional MBPC program aims at the voice service, which requires a higher grade of delay performance. As the data service grows and the number of users increases, the bit error rate (BET) is more sensitive to SINR. The low SINR level of MBPC cannot meet such QoS. From Fig. 4, it can be seen that the DCPC can improve the SINR level while reducing the user's transmit power, and the CDCPC program can significantly improve the user's SINR level. Compared to MBPC, CDCPC can save about 19% of the energy, and DCPC can save 13% of the energy.

As the cross-layer interference is more and more serious, D-layer of the transmission power in the bias scheme will be limited, and energy saving will be reflected in

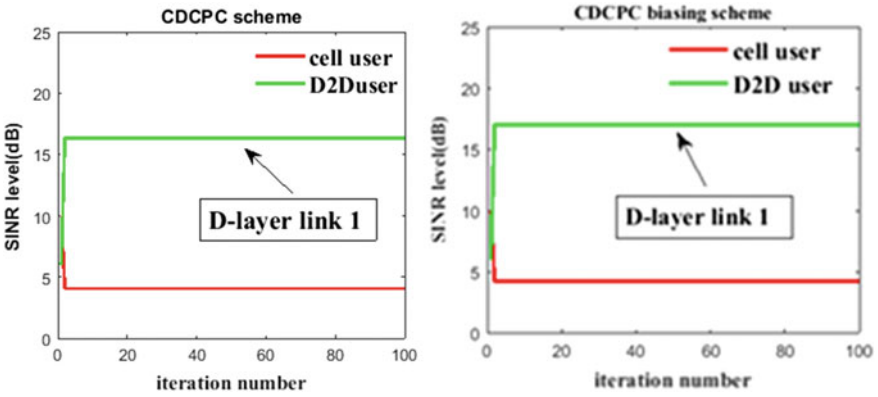


Fig. 5. Comparison of CDCPC and CDCPC programs after increasing bias.

the S-layer, which in turn will compensate for D-layer user channel conditions. It can be seen from Fig. 5, when using the bias scheme, the S-layer users can save about 66.7% energy than the original one.

5 Conclusion

Energy-efficient data communication has a huge impact on the user communication experience and is important for green communication. According to the people's expectations, D2D communication would significantly reduce the load pressure of the base stations and improve the system's spectrum utilization; but it also would bring the corresponding interference problem inevitably.

Based on the quality of service of users, this paper studies the power allocation strategy of D2D communication under the hybrid network topology of the single cell. And, we design a constraint scheme on the basis of predecessors. We also add a bias to the original scheme for further optimization, the simulation results can be seen by comparing the two schemes. The user's transmission power in S-layer is greatly reduced after increasing the bias program, which will further reduce the interference, increase the system capacity and achieve energy-saving purposes. It shows that our program is feasible.

Acknowledgments. This work was supported by Hainan Provincial Key R. & D. Projects of China (ZDYF2016010 and ZDYF2018012) and the National Natural Science Foundation of China (No. 61661018).

References

1. Despins, C., Labeau, F., Ngoc, T.L.: Leveraging green communications for carbon emission reductions: Techniques, testbeds, and emerging carbon footprint standards. IEEE Trans. Commun. Mag. **49**(8), 101–109 (2011)

2. Wu, J.S.: Green wireless communications: from concept to reality. IEEE Trans. Wirel. Commun. **19**(4), 4–5 (2012)
3. Zhu, J.K.: Capacity-power consumption and energy-efficiency evaluation of green wireless networks. IEEE Trans. Power Appar.Us Syst. **9**(2), 13–21 (2012)
4. Fodor, G., Dahlman, E., Mildh, G.: Design aspects of network assisted device-to-device communications. IEEE Trans. Commun. Mag. **50**(3), 170–177 (2012)
5. Lei, L., Zhong, Z., Lin, C.: Operator controlled device-to-device communications in LTE-advanced networks. IEEE Trans. Wirel. Commun. **19**(3), 96–104 (2012)
6. Gu, J., Bae, J.S., Choi, G.B., Min, C.Y.: Dynamic power control mechanism for interference coordination of device-to-device communication in cellular networks. In: IEEE Third International Conference on Ubiquitous and Future Networks, pp. 71–75 (2011)
7. Fodor, G., Reider, N.: A distributed power control scheme for cellular network assisted D2D communications. In: IEEE Global Telecommunications Conference, pp. 1–6 (2012)
8. Liu, P., Hu, C., Peng, T., Wang, W.: Admission and power control for Device-to-Device links with quality of service protection in spectrum sharing hybrid network. In: IEEE International Symposium on Personal Indoor and Mobile Radio Communications, pp. 1192–1197 (2012)
9. Yang, F.Y., Li, H.: Research on communication power control of cellular network terminal. China New Commun. **18**(21), 7–9 (2016)
10. Yates, R.D.: A framework for uplink power control in cellular radio systems. IEEE J. Sel. Areas Commun. **13**(7), 1341–1347 (2002)
11. Chi, S.W., Leung, K.K.: A generalized framework for distributed power control in wireless networks. IEEE Trans. Inf. Theory **51**(7), 2625–2635 (2005)

Research on User Grouping Strategy in NOMA Downlink Networks

Yun-Xiang Guo[1,2] and Hui Li[1,2(✉)]

[1] College of Information Science and Technology, Hainan University, Haikou, China
{ieguoyunxiang,hitlihuilll2}@163.com
[2] Marine Communication and Network Engineering Technology Research Center of Hainan Province, Hainan University, Haikou, China

Abstract. NOMA is a new type of multiple access technology proposed by NTT DOCOMO to improve spectral efficiency and system throughput and to be compatible with existing communication network architectures. In the NOMA downlink network, users in a cell are divided into several user groups, and the user grouping strategy has a large impact on the performance of the NOMA downlink network system. Therefore, in this paper, we study the existing user grouping algorithms in NOMA, compare their advantages and disadvantages, and propose a new user grouping algorithm based on greedy algorithm. Simulation results show that our algorithm performs well in reducing computational complexity and improving system performance.

Keywords: NOMA · User grouping strategy · Greedy algorithm

1 Introduction

In order to meet the demand for more user access and higher spectrum efficiency in next-generation mobile communications, ITU's requirements for 5G include: high capacity, low power consumption, wide coverage, low delay, and high reliability. However, the traditional multiple access technology cannot meet the above requirements of wireless communication, so we need a new multiple access scheme. NTT DOCOMO proposes a non-orthogonal multiple access (NOMA) scheme that utilizes the power domain, which has become one of 5G's alternatives.

NOMA actively introduces interference information at the transmitting end and uses non-orthogonal methods to send the signal. The users in a cell are divided into several user groups, and users in the same group use the same frequency. In the receiving end, the user signals are distinguished in the power domain through the serial interference cancelation technology.

Orthogonal multiple access is still used between different user groups, and different user grouping strategies have a significant impact on NOMA performance. Generally, users with high channel gains and users with low channel gains are selected as a group where users with low channel gains are allocated more power, and users with high channel gains are allocated less power. There are intra-group user interference and inter-group user interference in NOMA, and different user grouping strategies will

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 914–920, 2019.
https://doi.org/10.1007/978-981-13-6264-4_108

affect the above two kinds of interference in different degrees, which will affect system performance.

In addition, the key technologies in NOMA include the power-domain multiplexing at the transmitting end and the interference cancelation at the receiving end. This paper mainly studies the user grouping algorithm.

2 User Grouping Algorithm Applied to NOMA

2.1 Random User Grouping Algorithm

The random user grouping algorithm randomly selects users to implement user grouping in the candidate set and does not need to process information such as user channel state. Compared with other user grouping algorithms, the algorithm has the lowest complexity. However, this algorithm will cause large interference between users. Therefore, the system performance of the random user grouping algorithm is the worst.

In [1], all users in the cell are first randomly divided into a set of an equal number of users, and then the user with the best channel quality and the user with the worst channel quality are selected in each user set for grouping. Since each user set in the user grouping algorithm is obtained by random selection, there are cases where the quality of the user channel is generally better in some sets, and the quality of the user channel in other sets is generally poor. In this case, the channel gain difference of the users in the same group is too small, and the interference within the group is too large.

2.2 Exhaustive User Grouping Algorithm

In order to maximize the downlink capacity of the NOMA system, an exhaustive algorithm can be used. First, a full space search is performed on all possible user grouping results, and then the system performance under different user grouping results is compared to select the optimal performance.

Assuming that the number of users in the system is M and the number of superimposed users on subcarrier n is k_n, there are a total of $C_M^{k_n}$ user grouping modes on subcarrier n. It can be seen that when the system has N subcarriers, the complexity of the exhaustive algorithm is as high as $\left(C_M^{k_n}\right)^N$, so the algorithm is not applicable in practice.

2.3 Other User Grouping Algorithms

In [2–5], the user grouping algorithm based on the user channel gain considers the user channel state in the user grouping process, but the algorithm only uses the channel gain to group the users. Therefore, the grouping result can only guarantee to overcome the user interference in the group, and there will still be inter-group interference between different NOMA user groups.

In [6, 7], the user grouping algorithm using the user channel correlation can avoid certain inter-group interference and intra-group interference by utilizing the channel correlation and the channel gain difference. However, the algorithm needs to set the user channel correlation threshold. On the one hand, the user grouping result is closely

related to the threshold, but the threshold setting is random; on the other hand, the algorithm uses the correlation threshold to select the user group, which results in some user groups whose channel gain difference is large but the channel correlation is slightly smaller than the threshold, which limits the system performance.

3 User Grouping Algorithm Based on Greedy Algorithm

3.1 NOMA Downlink System Model and Optimization Problem

Assuming that the number of orthogonal subcarriers used for OFDM transmission is N, the total bandwidth of the system is B. The average power allocation is performed for the subcarriers, and the bandwidth of the single subcarrier is B/N. The power allocated by the i-th user on subcarrier n is denoted as $p_{i,n}$, and the transmitted signal on subcarrier n is denoted as $x_{i,n}$. The sum of the transmit powers of all users cannot be greater than the total transmits power P_{tot} of the system. Since the transmitting end of the NOMA system uses power multiplexing technology, multiple users share the same time-frequency resource unit. The number of users superimposed on the subcarrier n is represented by k_n. After k_n users are superimposed, the superimposed signal s of the base station on subcarrier n is expressed as

$$s_n = \sum_{i=1}^{k_n} \sqrt{p_{i,n}} x_{i,n} \tag{1}$$

Therefore, it can be obtained that the receiving signal $y_{k,n}$ of the receiving end on the subcarrier n can be expressed as

$$y_{k,n} = h_{k,n} \sum_{i=1}^{k_n} \sqrt{p_{i,n}} x_{i,n} + w_{k,n} \tag{2}$$

where $h_{k,n}$ and $w_{k,n}$ represent the channel gain and additive Gaussian noise of the base station to the receiving end on the subcarrier n. The receiving end not only contains its own useful signals, but also contains interference signals of other users. We can use SIC for signal detection. After SIC detection processing, the SINR of user m on subcarrier n is

$$SINR_{m,n} = \frac{p_{m,n}\beta_{m,n}}{1 + \sum_{i=1}^{m-1} p_{i,n}\beta_{m,n}} \tag{3}$$

where $\beta_{m,n} = |h_{m.n}|^2/\sigma_n^2$ denotes carrier to interference noise ratio.

It can be seen that SINR is related to the actual channel conditions and power of each user. After SIC detection processing, throughput of user m on subcarrier n is

$$R_{m,n} = \frac{B}{N}\log_2\left(1 + SINR_{m,n}\right) \tag{4}$$

and the total capacity of subcarrier n is

$$R_n = \frac{B}{N}\sum_{k=1}^{K} s_{k,n}\log_2\left(1 + SINR_{k,n}\right) \tag{5}$$

It can be seen that the throughput of the user on a single subcarrier is closely related to the manner of user grouping on the subcarrier and the transmit power of the superimposed user. In order to maximize the throughput of the NOMA system, the user grouping strategy and the power allocation algorithm can be translated into the following optimization problems

$$\begin{aligned} &\max \sum_{n=1}^{N}\sum_{k=1}^{K} s_{k,n}R_{k,n} \\ &s.t.\begin{cases} C_1 : p_{k,n} \geq 0 \\ C_2 : \sum_{n=1}^{N}\sum_{k=1}^{K} p_{k,n} \leq P_{tot} \\ C_3 : s_{k,n} \in \{0,1\} \\ C_4 : \sum_{k=1}^{K} s_{k,n} = k_n \end{cases} \end{aligned} \tag{6}$$

where C_1 represents that the user power on the n-th subcarrier is not less than 0, and C_2 represents that the sum of power on all subcarriers is not greater than P_{tot}. $s_{k,n}$ represents whether the k-th user occupies the n-th subcarrier, and if the k-th user occupies the n-th subcarrier, sets $s_{k,n}$ to 1; otherwise, sets $s_{k,n}$ to 0.

3.2 User Grouping Strategy and Algorithm Flow

The above optimization problem is a joint optimization problem that combines user grouping and power allocation on subcarriers. Obviously, there is a high complexity when solving the global optimal solution through joint optimization. In order to reduce the complexity of resource allocation, we consider the user grouping problem on subcarriers separately.

Considering that the greedy algorithm has lower complexity, it can be applied to the user grouping strategy on the subcarriers in the NOMA system. The basic idea of the greedy algorithm is that when solving the problem domain, the best choice in the current step is always taken as the optimal strategy. It can be seen that the greedy algorithm does not start from the global optimal point of view and belongs to the local optimal solution. Assume that the subcarriers are equally allocated with power, and the power allocation of the superimposed users on each subcarrier adopts the FTPA algorithm. The specific algorithm flow of the user grouping is as follows:

Step1: Initialize each parameters: the superimposed user set is Ω_n on each subcarrier, the subcarrier set $A = \{1, 2, \ldots, N\}$, and the user set $U = \{1, 2, \ldots, K\}$, let $s_{k,n} = 0$, the rate of user k is $v_k = 0$, where subcarrier $n \in A$, user $k \in U$, the number of iterations $i = 1$.
Step2: For the i-th subcarrier, if the capacity of two or more users in the user set U is zero, the user u with the best channel gain is selected from among the users; otherwise, the user u with the largest $R_{k,i}/v_k$ in the user set U is selected. Group user u into Ω_i, remove it from the user set U and update v_k.
Step3: Find other superimposed users on the i-th subcarrier: A user j in the user set U is arbitrarily selected, and the total capacity of the user j and all the users in the set Ω_i on the i-th subcarrier is calculated by (5), and the user j that takes the maximum value of the total capacity is found as the user J. Group user J into Ω_i, remove it from the user set U and update v_k.
Step4: If the number of elements in the set Ω_i is less than k_i, return to Step3; otherwise, continue to the next step.
Step5: The grouping of the superimposed users on the i-th subcarrier is completed, $i = i + 1$, and let $U = \{1, 2, \ldots, K\}$.
Step6: If $i \leq N$, return to Step 2; otherwise, jump out of the iteration, the user grouping on the subcarrier is completed.

The complexity analysis of the user grouping based on the greedy algorithm on a single subcarrier is performed, and the analysis process considers the number of user combinations on the subcarriers. Assuming that the number of users is superimposed on subcarrier n as k_n, and the total number of users that can be selected is N, there are $C_N^1 + C_{N-1}^1 + \cdots + C_{N-k_n+1}^1 = 1/2 * (2N - k_n + 1)k_n$ user combinations on subcarrier n. Compared with the exhaustive algorithm, the computational complexity is effectively reduced.

4 Simulation Results

In order to verify the system performance of the proposed algorithm, the simulation verification method in [8] is used to perform simulation analysis on MATLAB. The simulation scenario is set to a single cell of a single base station, and the location of the

Table 1. Parameters used in simulation

Quantity	Value	Unit
Cell radius	500	m
Base station transmit power	43	dBm
System bandwidth	4.32	MHz
Receiver noise power spectral density	−169	dBm/Hz
Path loss model	128.1 + 37.6lg(r),	dB
Number of base station transmit antennas	2	–

users in the network are randomly and evenly distributed. The simulation parameters are shown in Table 1.

Figures 1 and 2 show the system capacity obtained by using greedy algorithm, exhaustive algorithm, and random algorithm when the number of users is 4, 8, 12, 16, 32, 48, and 64, respectively. Random algorithm randomly selects two users in a cell into a group, the complexity is low, but the system capacity decreases due to large interference.

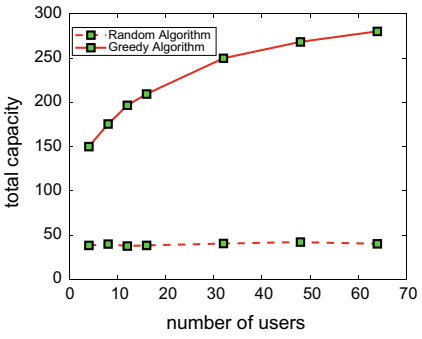


Fig. 1. Greedy algorithm vs Random algorithm

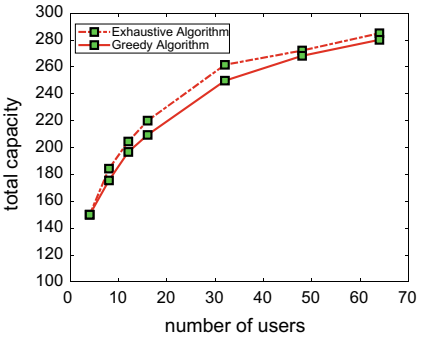


Fig. 2. Greedy algorithm vs Exhaustive algorithm

nce caused by random grouping. Therefore, the random algorithm system capacity is lower than 50 Mb/s in this simulation scenario.

The exhaustive algorithm can maximize the total capacity, but it is necessary to exhaust all user grouping results in the user grouping, which has higher complexity, and the system capacity increases as the number of users increases. The algorithm proposed in this paper has a total system capacity similar to that of the exhaustive algorithm. When the number of users is 64, the system capacity of this algorithm reaches 280.2 Mb/s, which is very close to the system capacity of the exhaustive algorithm at this time of 285.1 Mb/s. But its complexity is lower than the exhaustive algorithm.

5 Conclusions

This paper first introduces the meaning of user grouping in the NOMA system, and then introduces two user grouping schemes, random algorithm, and exhaustive algorithm. However, these two algorithms have their own disadvantages such as high computational complexity and excessive randomness of grouping results. Since the greedy algorithm can effectively reduce the computational complexity under the

premise of similar performance with the exhaustive algorithm, we propose a user grouping scheme based on greedy algorithm. The simulation results show that the algorithm mentioned in this paper greatly improved the system capacity compared with random algorithm, and in the case where the complexity is lower than exhaustive algorithm, the system capacity is close to the system capacity when exhaustive algorithm is used.

Acknowledgments. This work was supported by Hainan Provincial Key R. & D. Projects of China (ZDYF2016010 and ZDYF2018012) and the National Natural Science Foundation of China (No. 61661018).

References

1. Eddo, Z., Hojeij, M.R., Nour, C.A., et al.: Evaluation of intra-subband power allocation for a downlink non-orthogonal multiple access (NOMA) system. In: 2016 IEEE Globecom Workshops (GC Wkshps), pp. 1–7. IEEE Press, Washington (2016)
2. Saito, Y., Kishiyama, Y., Benjebbour, A., et al.: Non-orthogonal multiple access (NOMA) for cellular future radio access. In: 2013 IEEE 77th Vehicular Technology Conference (VTC Spring), pp. 1–5. IEEE Press, Dresden (2013)
3. Zhang, X., Gao, Q., Gong, C., et al.: User grouping and power allocation for NOMA visible light communication multi-cell networks. IEEE Commun. Lett. **10**(99), 1–4 (2016)
4. Chen, S., Peng, K., Jin, H.: A suboptimal scheme for uplink NOMA in 5G systems. In: 2015 International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 1429–1434. IEEE Press, Dubrovnik (2015)
5. Ali, M.S., Tabassum, H., Hossain, E.: Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems. IEEE Access **4**(12), 6325–6343 (2016)
6. Zhang, H., Zhang, D.K., Meng, W.X., et al.: User pairing algorithm with SIC in non-orthogonal multiple access system. In: 2016 IEEE International Conference on Communications (ICC), pp. 1–6. IEEE Press, Kuala Lumpur (2016)
7. Ali, M.S., Hossain, E., Kim, D.I.: Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: user clustering, beamforming and power allocation. IEEE Access **5** (99), 1–13 (2017)
8. Liu, S., Zhang, C., Liu, G.M.: User selection and power schedule for downlink non-orthogonal multiple access (NOMA) system. In: Proceedings of 2015 Communication Workshop (ICCW), pp. 2561–2565. IEEE, London (2015)

Ad Hoc Networks

An Evolving Scale-Free Networking Model for Vehicular Ad Hoc Network

Xinyu Qu, Erwu Liu(✉), and Rui Wang

School of Electronics and Information Engineering,
Tongji University, Shanghai, China
erwuliu@tongji.edu.cn

Abstract. Recent years have witnessed extensive research in the characteristics of vehicular ad hoc network (VANET) topology. Some papers in the literature use a simple model to generate a VANET topology at a certain time. Different from the existing works, we consider the effect of node addition, node deletion and link loss due to node mobility and keep the network evolving. We employ realistic vehicular traces instead of traffic model or simulator and propose an evolving networking model for VANET. We find that the evolved VANET exhibits a scale-invariant feature under certain conditions. In a word, the network has a power-law degree distribution and a good connectivity and survivability to deliberate node attacks and random node failures. Our work will help establish a strongly connected VANET topology.

Keywords: VANET · Realistic traces · Connectivity · Survivability

1 Introduction

Vehicular ad hoc network (VANET) has recently gained wide attention in both academia and industry. It is provided to support various services for Intelligent Transportation Systems (ITS) such as dynamic route planning [1] and safety message propagation [2]. Defined in [3], the Dedicated Short Range Communications (DSRC) technology helps improve the traffic safety and efficiency, by enabling direct vehicle-to-vehicle communications.

Recently, the topology characteristics of VANETs have attracted the attention of many researchers. They tend to use or design traffic model including mobility model and road structure to simplify the study of VANETs [4,5]. However, realistic road structure and traffic mobility are much more complicated and simple changes on widely used models still fail to represent the real situation. Hence in recent years we see the tendency to use realistic dataset in VANET studies and analyze the network topology by complex network theory. Some of the works generate the topology by unit disc graph (UDG) model and observe interesting characteristics such as the scale-free feature and small-world effect [6,7]. Some literatures also argue that VANETs are not scale-free networks [8,9]. These research use very simple models and the results only depends on the node

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 923–930, 2019.
https://doi.org/10.1007/978-981-13-6264-4_109

density and distribution of the dataset. In this paper we will consider the interactions between vehicles and model VANET by complex network theory. Since a scale-free network has a good connectivity and survivability [10], in this paper we will also try to figure out processes to make the network evolve to a stable scale-free one and improve the realistic VANET.

2 Dataset and Metrics

We will use complex network theory to evolve and analyze the network. In this section we will introduce the dataset we use and the definitions of some network metrics. The dataset we employ is not generated by simulator or on the basis of realistic data like [9]. All vehicular mobility data are truly collected from taxis and buses in a certain city.

2.1 Vehicular GPS DataSet

We collect two sets of realistic vehicular GPS traces. One consists of over 4000 taxis and buses in Shanghai which is the same one in [8] and another contains over 7000 taxis in Shenzhen. We employ the vehicles ID and the location information which includes the latitude, longitude and the recording time of each vehicle. The start and end time of record are different for each vehicle and we have vehicles without records in some time periods. To deal with it, we set: (a) if we do not find more than five records whose time difference to a certain time is within 3 min, it means that the vehicle has no location information at that time; (b) if we get enough data, we then use interpolation algorithm to calculate the latitude and longitude at the certain moment.

2.2 Network Metrics

We define $G(\mathbb{V}(t), \mathbb{E}(t))$ as the instantaneous graph at time t. $\mathbb{V}(t) = \{v_i\}$ denotes the set of the nodes and $\mathbb{E}(t) = \{e_{ij}(t) | v_i, v_j \in \mathbb{V}(t), i \neq j\}$ is the set of edges modeling the links between each node. The number of edges of the node represents the degree of the node.

In the Unit Disc Graph (UDG) model [11], there is an edge between nodes v_i and v_j if the distance between these two nodes is not bigger than the communication range r. The nodes within one node's range are called its neighbors. $\mathcal{L}_i$ is a set which contains the neighbors of node v_i. In this paper, we use UDG as a baseline model for topology construction.

The Connected Component (CC) is a non-empty subgraph of network at time t. In such subgraph, nodes communicate with each other, which means there is always a path that one node can reach another. The Greatest Connected Component (GCC) is the one which contains the most number of nodes in the network.

3 VANET Evolution Process

In this section, we present the mechanisms involved in VANET evolution. Since VANET contains vehicles driven by human beings, we believe that links between vehicles are not totally random and obey a rule in social network theory. Hence we use complex network theory to evolve VANET. Similar to [9], our analysis aims at the physical topology of VANET and is independent to communication technology and protocol. Given the size of our dataset and the complexity of our proposed processes, we will not consider obstacles and complicated channel models in the network because they require expensive computations at each time of every node and do not scale well to our scenario. We assume that each vehicle node is able to detect whether it has lost a link or joins the network and obtains the degrees of the nodes within its communication range to perform related processes.

3.1 Node Addition and Deletion

If a taxi comes into the network, it will find all the disconnected nodes within its communication range r and do the preferential attachment to generate m new links. The probability that it choose node i is

$$\prod(v_i) = \frac{k_i}{\sum\limits_{v_j \in \mathcal{L}_i} k_j}, \tag{1}$$

where k_i is the degree of node v_i.

We rule that if a vehicle leaves the network, all the edges attached to it should be removed. Hence it is easy to make the network very low-connected during the evolution. To ensure the connectivity, a common way is using link compensation to avoid fragmenting the network into several isolated clusters [12]. We assume vehicles that lost edges because of this event has the probability p_c to do link compensation. The probability that it chooses node v_i is determined by Eq. (1).

3.2 Event Caused by Node Mobility

Since node mobility is very strong in VANET, it is easy to have connection loss during the network evolution. Hence the topology would not be a stable one and we will have link deletion and compensation frequently. We will have different situations here. If a node moves and the previously connected nodes are still within the range, then we will perform no process on such node. For an isolated node, which means it is not connected to any other nodes in the network, we denote the previous moment as time T_p and the current one as time T_c. If it was isolated at time T_p and it is still isolated at time T_c, we will perform no process on such node. However, it may detect some other nodes in its communication range at time T_c. If this occurs, we regard it the same as node addition and do the processes accordingly. If one node has lost edges and it finds no disconnected

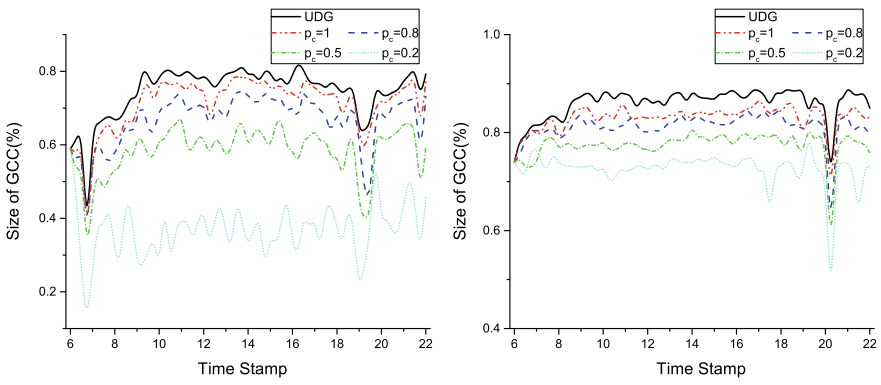
nodes in its range, then it will perform no process but the other nodes which were attached to it at time T_p will do the compensation process. On the other hand, if the node has lost edges and it finds some disconnected nodes in its range, then it will do the link compensation here.

4 Simulation Experiments

In this section, we evolve the network with realistic vehicular traces and show the results. We use UDG model as a baseline model for topology construction, and compare the sizes of GCC at each time generated by the UDG model with the ones by our proposed algorithm. Then we show the degree distribution of the generated network and see how it evolves to be a power-law one over time. Finally, we present the survivability of different topology with various p_c.

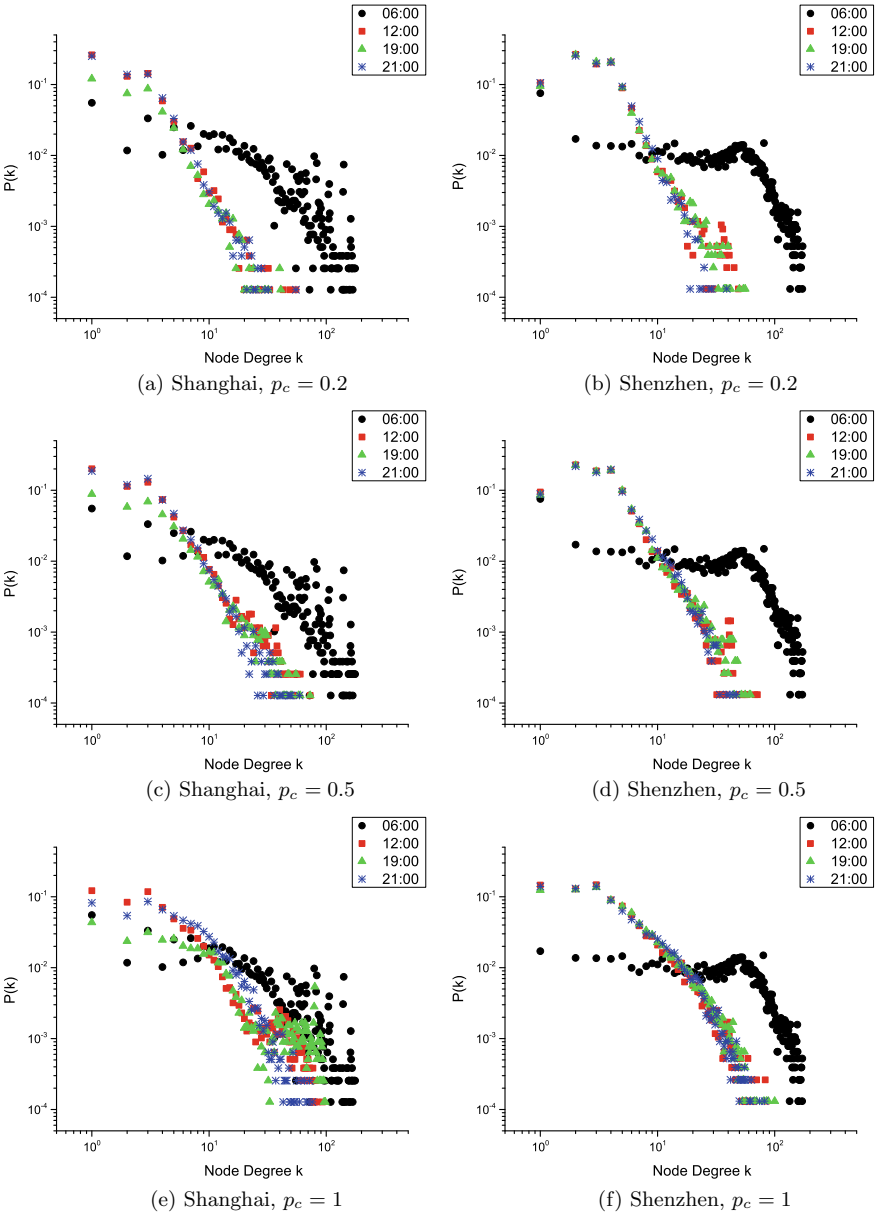
4.1 Size of GCC

Figure 1 shows the sizes of the GCC by our evolution process under different p_c from 6:00 to 22:00. From Fig. 1, we find that in Shanghai if $p_c < 0.5$, it is hard to make the network connected. The negative impact of node deletion and connection loss gets more serious because the link compensation here only plays a minor role. When $p_c \geq 0.5$, the loss in the size of the GCC by applying the mentioned mechanisms is mostly less than 15%. Since we can generally obtain the GCC with about 60% of the nodes in Shanghai, it is proper to set $p_c \geq 0.5$. We can also see that the GCC is very small at around 6:30 and 19:00. The reason is that taxies account for a large proportion of the vehicles in the dataset and 6:30 and 19:00 are the shift time for taxi drivers to change off in Shanghai. Hence there are few nodes and the density is relatively low. In Fig. 1, the negative impact of node deletion and connection loss in Shenzhen when $p_c < 0.5$ is less



(a) Shanghai (b) Shenzhen

Fig. 1. Size of GCC under different p_c from 6:00 to 22:00



(a) Shanghai, $p_c = 0.2$ (b) Shenzhen, $p_c = 0.2$

(c) Shanghai, $p_c = 0.5$ (d) Shenzhen, $p_c = 0.5$

(e) Shanghai, $p_c = 1$ (f) Shenzhen, $p_c = 1$

Fig. 2. Degree distribution at different times under various p_c

serious than the one in Shanghai. And the sizes of GCC with different p_c are comparatively smaller which is less than 10%. The change shift in the early morning is less obvious in the simulation and the evening shift time for taxi drivers is also different which is about 20:30.

4.2 Degree Distribution

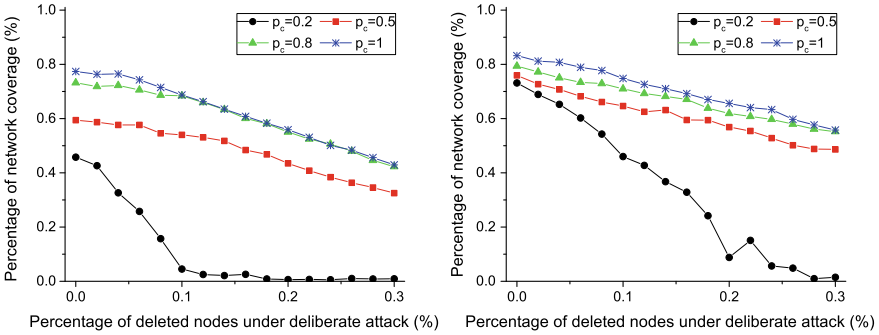
Here we give the degree distribution of the VANET at 6:00, 12:00, 19:00, and 22:00 in Shanghai and Shenzhen. In Fig. 2, we see that the networks of both cities are more likely scale-free when $p_c \leq 0.5$. Our proposed network evolution processes work well in both scenarios. And compared to the results in [8], we successfully evolve the VANET into a scale-free one with the same dataset of Shanghai. When $p_c < 0.5$, the number of links in the network decreases over time, which tends to weaken the network connectivity. When $p_c > 0.5$, both the node addition and link compensation mechanisms increase the number of links in the network during the network evolution. The longer vehicle nodes stay in the network, the more likely they will connect to all the nodes within the range. Therefore the evolving network will be similar to the one generated with the UDG model as time increases. Thus the both the degree distributions of the evolving VANETs with $p_c = 0.8$ and $p_c = 1$ does not obey power law.

4.3 Survivability

To evaluate the survivability, we use the metric proposed in [6], which is denoted as

$$C(T) = \sum_{i=1}^{N} k_i(T)/[N(N-1)], T \geq 1, \tag{2}$$

where T is the maximum number of hops between nodes, N is the network size, $k_i(T)$ is the number of nodes node i can reach within T hops and $C(T)$ is the network coverage. The survivability of the network in Shanghai and Shenzhen under deliberate attacks and random failures are shown in Figs. 3 and 4, respectively. Here we take as input the network data at 22:00 when the network has reached the steady state. We find that the reduction in percentage of network coverage is relatively small when 30% of nodes are deleted under random failures and when $p_c \geq 0.5$, the network survivability to deliberate attacks is also dramatic.



(a) Shanghai (b) Shenzhen

Fig. 3. Survivability to deliberate attacks

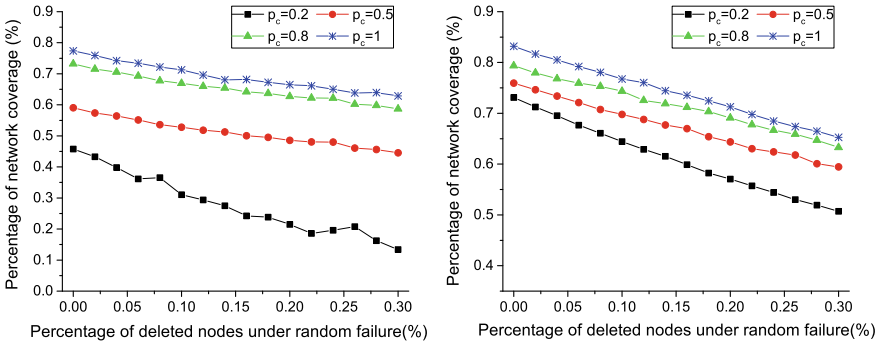


Fig. 4. Survivability to random failures

Since link compensation has little influence on the network when $p_c = 0.2$, it leads to the worst performance as shown in the figures. We can also find that the survivability is even better in Shenzhen scenario. The reason is that there are more vehicles in the dataset of Shenzhen and the nodes can find more neighbors to compensate the loss.

5 Conclusion

In this study, we propose an evolving model for VANET and analyze the topology characterized by GPS traces. We use node addition, node deletion and link loss to model the influence brought by node mobility. Based on this, we include preferential attachment and link compensation mechanisms during the evolution and observe that we can ensure a good network connectivity and survivability. We also find that the scale-invariant feature will emerge under certain conditions.

Though we have proposed a promising approach to evolve the VANET topology, we did not use gain insight into such phenomenon and discuss how a VANET would evolve and what it would turn into. And we did not consider other important aspects such as the traffic regulations and the presence of obstacles. In future work, we will address these issues in both the urban and rural scenarios.

References

1. Nzouonta, J., Rajgure, N., Wang, G., Borcea, C.: Vanet routing on city roads using real-time vehicular traffic information. IEEE Trans. Veh. Technol. **58**(7), 3609–3626 (2009)
2. Torrent-Moreno, M., Mittag, J., Santi, P., Hartenstein, H.: Vehicle-to-vehicle communication: fair transmit power control for safety-critical information. IEEE Trans. Veh. Technol. **58**(7), 3684–3703 (2009)
3. Commission, F.C., et al.: Fcc 99-305. FCC Report and Order (1999)

4. Khabazian, M., Ali, M.K.M.: A performance modeling of connectivity in vehicular ad hoc networks. IEEE Trans. Veh. Technol. **57**(4), 2440–2450 (2008)
5. Viriyasitavat, W., Bai, F., Tonguz, O.K.: Dynamics of network connectivity in urban vehicular networks. IEEE J. Onselected Areas Commun. **29**(3), 515–533 (2011)
6. Zhu, Y., Chen, C., Gao, M.: An evaluation of vehicular networks with real vehicular GPS traces. EURASIP J. Wirel. Commun. Netw. **2013**(1), 190 (2013)
7. Zhang H., Li, J.: Topology analysis of VANET based on complex network. In: LISS 2014, pp. 1143–1148. Springer, Heidelberg (2015)
8. Liu, X., Li, Z., Li, W., Lu, S., Wang, X., Chen, D.: Exploring social properties in vehicular ad hoc networks. In: Proceedings of the Fourth Asia-Pacific Symposium on Internetware, no. 24, pp. 1–7. ACM, Qingdao (2012)
9. Naboulsi, D., Fiore, M.: On the instantaneous topology of a large-scale urban vehicular network: the cologne case. In: Proceedings of the Fourteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing, pp. 167–176. ACM, Bangalore (2013)
10. Albert, R., Jeong, H., Barabási, A.-L.: Error and attack tolerance of complex networks. Nature **406**, 378–382 (2000)
11. Clark, B.N., Colbourn, C.J., Johnson, D.S.: Unit disk graphs. Discret. Math. **86**(1), 165–177 (1990)
12. Wang, Y., Liu, E., Jian, Y., Zhang, Z., Zheng, X., Wang, R., et al.: NLL: a complex network model with compensation for enhanced connectivity. IEEE Commun. Lett. **17**(9), 1856–1859 (2013)

An Energy-Efficient Sensor Protocol for Information via Negotiation

Xiangdong Yi(✉), Jiasong Mu, and Xiang Liu

Springer-Verlag, Computer Science Editorial, Tiergartenstr. 17,
69121 Heidelberg, Germany
yixiangdong52@qq.com

Abstract. As a product of the combination of computing, communication and sensor technologies, wireless sensor networks become an hot research branch at the area of wireless communications. In wireless sensor networks, energy efficiency is the most important factors affecting performance of sensor networks. Excellent routing algorithms could significantly increase the utilization of energy and prolong service life of sensor network. Since data-centric routing algorithms are easy to deploy and have stronger robustness in applications [1], many new protocols are developed on the basis of it. This article analyzes some commonly used symbolic algorithms, describes their advantages and disadvantages, and improves on the basis of the SPIN protocol to improve the energy efficiency and extend the network lifetime.

Keywords: Wireless sensor networks · Data-centric routing · Data fusion SPIN

1 Introduction

Wireless sensor network is particularly suitable for highly dynamic and malicious areas. It consists of hundreds of thousands of sensor nodes with computing and communication functions. The sensor node collects, processes, and transmits data acquired from the physical environment to external base stations. Currently, wireless sensor networks have entered various applications and systems, and their needs and characteristics are very different. They are mainly used for disaster relief, emergency rescue operations, military, human settlements monitoring, medical care, radioactivity monitoring, etc [2].

Traditional routing protocols generally use addresses as the basis for node identification and routing [3]. In wireless sensor networks, a great quantity of nodes are randomly distribution in the environment. The focus is on the moni-

Please note that the LNCS Editorial assumes that all authors have used the western naming convention, with given names preceding surnames. This determines the structure of the names in the running heads and the author index.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 931–938, 2019.
https://doi.org/10.1007/978-981-13-6264-4_110

toring area's sensing data, rather than the specific node's acquired information, and because of wireless transmission. Moreover, due to the high density and overhead of WSN, it is not real to allocation global ID to each node [4]. Thus, addressing based routing is not feasible for application in Wireless Sensor Networks. For the sake of solving this problem, data-centric routing was proposed.

In a data-centric protocol, when the sensor sends its data to the sink, the intermediate sensor can perform some form of aggregation of data from some sensors and send the aggregated data to the sink [5]. The core technology of data-centric routing is data aggregation. In data fusion, nodes can not only simplify the data, but also can synthesize the data generated by multiple sensor nodes into meaningful information according to the characteristics of the data for a specific application environment, thereby improving the accuracy of the sensing information. The robustness of the system is enhanced, redundancy is eliminated, power consumption is reduced, and sensor node lifetime is extended [6]. Typical data-centric routing protocols include flooding and gossiping, sensor protocol for information via negotiation, directed diffusion, and so on.

2 Data-Centric Routing

2.1 Flooding and Gossiping

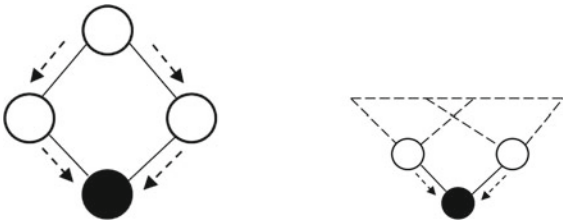
Flooding is a traditional mechanism for transmitting data in a sensor network and is an old and simple algorithm. Its working principle is as follows: Each sensor in the network receives a data packet and sends the accepted packet to all its neighbors in the form of a broadcast. And repeats this course until all nodes in the network receive the data packet [7,8]. Its priority is easy to implement, and nodes do not need to keep information about their neighbors and use complex information route discovery algorithms and topology maintenance, but implosion problems, overlap problems, and resource blindness are inherent deficiencies.

Implosion: As shown in Fig. 1a, a node may receive the same data copy from multiple nodes, so the resource is wasted due to extra sending and receiving operations [9].

Overlapping: The two nodes may sense a common area or the coverage areas overlap, each sensor sends the sensed data to their public neighbor, so that two similar data is transmitted to the neighbor nodes. It is shown in Fig. 1b.

Resource blindness: The node's available energy is not considered in this algorithm, thus resulting in many redundant transmissions.

Gossip is an updated version of the flooding. Once a node obtains this packet, unlike flooding, the packet is not broadcast. Instead, it randomly selects a neighbor and forwards the packet to it. When a neighbor node obtains packet, it repeats this trend until the packet arrives at its whither [10]. Compared with flooding, avoiding information explosion problems and low energy consumption, but still cannot solve a part of the overlap phenomenon and the blind use of resources, and the average delay of data transmission lengthens, the transmission speed slows down [11].



(a) Implosion problem

(b) Overlap problem

Fig. 1. Some samples from Implosion and Overlap problem.

2.2 Sensor Protocols for Information via Negotiation

SPIN is an information-propagation protocols based on negotiation and having energy adaptation capabilities. It is the first data center routing mechanism. Before the data is transmitted, the sensor nodes negotiate with each other and the negotiation system ensures the transmission of useful data. Nodes negotiate by sending metadata (i.e., data describing the data attributes collected by the sensor nodes) rather than the entire data collected. To avoid blindly using resources, all sensor nodes must monitor their respective energy changes. Before transmitting or receiving data, each node must check its own available energy status. If it is at a low energy level, some operations must be terminated (such as data forwarding operations).

SPIN uses three types of data packets: ADV, REQ, and DATA.

ADV is used for data broadcasting. When a node has data to send, it can broadcast with ADV packet containing metadata.

REQ is used to request sending data. When a node intends to receive DATA packets, it sends REQ packets.

DATA contains a packet of data collected by the sensor attached to the metadata header.

Before sending a DATA packet, the sensor node first broadcasts the ADV packet externally. If a neighbor is willing to receive the DATA packet after receiving the ADV packet, A REQ packet is sent to the node, and then the node sends a DATA packet to the neighbor. Similarly, DATA packets can be transmitted to sink nodes or base stations [12].

SPIN protocol based on data description negotiation mechanism and energy adaptation mechanism can well solve the problem of implosion and overlap in traditional flooding protocols, and effectively save energy and prolong the life cycle of the network. However, since it has no route optimization capability, there is no measure in the choice of path. The data broadcast mechanism of the SPIN protocol cannot ensure reliable transmission of data. If the node that needs data is far away from the source node and the node between the sensor

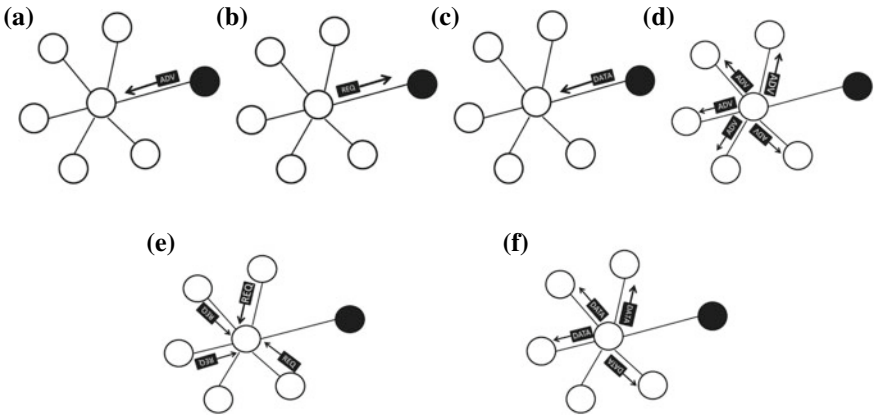


Fig. 2. SPIN protocol phases.

and the destination node does not pass in the data, the data will be lost and will not be transmitted to the destination node. Therefore, SPIN is not a good choice in applications such as intrusion detection because intrusion detection requires regular and reliable transmission of data packets (Fig. 2).

3 An Energy-Efficient Sensor Protocols for Information via Negotiation (EF-SPIN)

The SPIN protocol uses metadata for negotiation, reducing the transmission of redundant data in the network and avoiding the blind use of resources. However, the SPIN protocol also has the problems of blind forwarding and data inaccessibility. The former will waste energy, while the latter will cause network information not to be collected. These two problems have a very serious impact on wireless sensor networks. Therefore, aiming at the problems existing in the SPIN protocol, aiming at improving the utilization of energy, an improved wireless sensor network routing protocol EF-SPIN is designed. Add the flag F and the energy value E in the REQ message. The flag bit indicates whether the neighbor node owns the data. Regardless of whether the neighbor node owns the data, the REQ message is returned as long as the energy value is not lower than the set threshold. When there is new data to be forwarded by the node, it is forwarded according to the flag bit and the energy value. After the time threshold is exceeded, the node deletes the neighbor node that did not send the REQ response message from the neighbor list.

The protocol execution steps are described as follows (Fig. 3):

(1) The data source node broadcasts the ADV message and starts the timer T, giving the threshold TM a wait.

(2) The neighboring node receiving the ADV message decides the working mode according to its energy value. If the neighbor node energy value is lower than the set threshold, it will not respond, otherwise it will respond with its energy and whether it has the broadcasted data.
(3) When T is less than or equal to TM, the source node receives the REQ response packet to update its corresponding neighbor node energy E and flag F, and the source node selects one neighbor node with high energy from the neighbor nodes with Flag = 0. If the node energy is the same, a neighbor node is randomly selected, and data is sent to it to determine whether the neighbor node is a sink node. If yes, the data transmission ends, and step 5 is returned. Otherwise, the node is used as a data source node, and the procedure returns to step 1 to continue execution.
(4) When T is more than the TM, if the source node receives a REQ message with the flag bit set to 1, the source node removes the neighbor node that did not send the REQ response packet from the neighbor list, and then selects a large energy value in the neighbor list. Node, forcing data forwarding. Determine if this node is a sink node. If yes, return to step 5, otherwise, use the node as a data source node, return to step 1 to continue execution.
(5) Sink node receives the required data and the program terminates.

The following figure shows the working process of sending and receiving data in the EF-SPIN protocol based on the EF-SPIN protocol. There is no "data unreachable" problem within a specified time, and the energy with a search flag of 0 is large. The node forwards the data. If a "data unreachable" problem occurs after the timeout, the node with the highest energy value is forced to forward the data, and the unresponsive node is deleted at the same time. After the receiving node receives the data, the data transceiving process between the sending node and the receiving node ends. If the receiving node is not a Sink node, its role becomes the sending node, and this process is repeated again until the data is sent to the target node. In the EF-SPIN protocol, the source node will do the corresponding work according to whether the REQ message and the flag in the REQ message are received. Deleting nodes that do not respond to the time threshold value from the neighbor list can save the energy of the sensor nodes and increase the storage space of the sensor nodes. According to the energy priority principle, the EF-SPIN protocol determines that a neighbor node uses a single path to forward data, solves the "blind forwarding" problem, and solves the problem of "data unreachability" with forced forwarding. However, EF-SPIN also has certain problems. For example, mandatory forwarding may cause "routing loop" problems.

4 Comparison of Simulation Results

For wireless sensor networks, there is currently no unified standard to evaluate different routing protocols. According to the actual situation, this paper evaluates the EF-SPIN protocol with energy consumption at the same time and

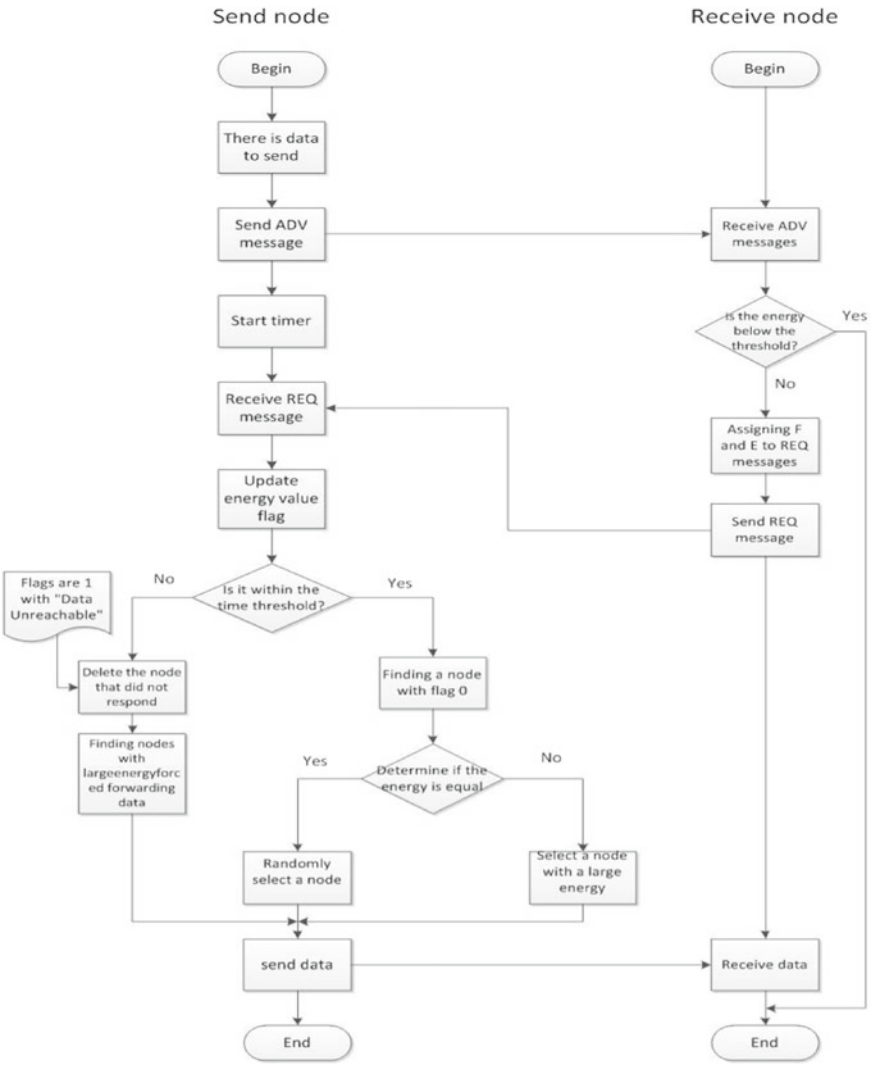
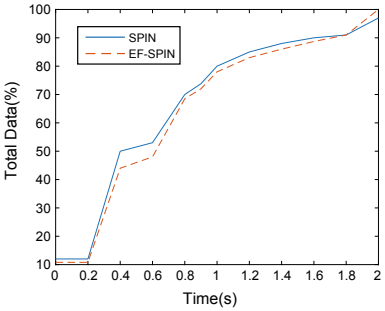
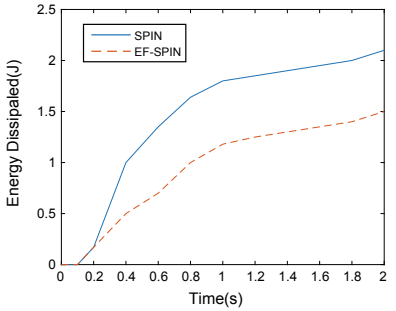


Fig. 3. Schematic diagram of the transmission process of the sending and receiving nodes of the EF-SPIN protocol.

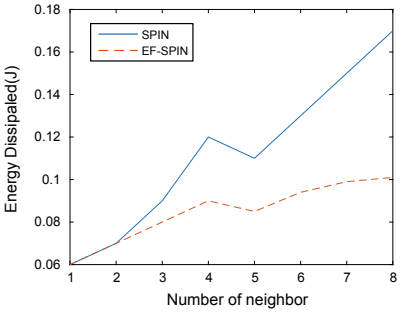
the total number of data packets received by Sink nodes. In order to reduce the errors in the simulation results, the results of the imitation in this paper are compared with the average value of the protocol executed 10 times. In this paper, a convergent node is set at each end of the rectangular area. The SPIN protocol does not have the problem of "data unreachable at the initial stage of the network. The following are the simulation results of this experiment (Fig. 4).



(a) The total number of data reaches the aggregation node and the corresponding time



(b) The total energy consumed by the system and the corresponding time



(c) Comparison of node energy consumption and number of neighbor nodes

Fig. 4. Comparison between SPIN and EF-SPIN simulation.

5 Conclusion

The routing method in WSN is a hot and emerging field of research, and the research results are limited but rapidly growing. In this article, the EF-SPIN protocol solves the "blind forwarding" and "data inaccessibility" problems of the SPIN protocol. On this basis, it still needs to continue to improve and research. In order to minimize power consumption and extend the service life of the network, it does not affect the data transmission is the common goal of wireless sensor network routing methods, and the wireless sensor network still faces many challenges.

Acknowledgments. This paper is supported by Natural Science Foundation of China (61271411), Natural Youth Science Foundation of China (61501326, 61401310). It also supported by Tianjin Research Program of Application Foundation and Advanced Technology (15JCZDJC31500), and Tianjin Science Foundation (16JCYBJC16500).

Appendix: Springer-Author Discount

LNCS authors are entitled to a 33.3% discount off all Springer publications. Before placing an order, the author should send an email, giving full details of his or her Springer publication, to http://orders-HD-individuals@springer.com to obtain a so-called token. This token is a number, which must be entered when placing an order via the Internet, in order to obtain the discount.

6 Checklist of Items to be Sent to Volume Editors

Here is a checklist of everything the volume editor requires from you:

- ☐ The final LaTeX source files
- ☐ A final PDF file
- ☐ A copyright form, signed by one author on behalf of all of the authors of the paper.
- ☐ A readme giving the name and email address of the corresponding author.

References

1. Akyildiz, I.F., Vuran, M.C.: Wirless Sensor Networks, vol. 4. Wiley (2010)
2. Dargie, W., Poellabauer, C.: Fundamental of Wireless Sensor Networks, pp. 168–175. Wiley (2010)
3. Akkaya, K., Younis, M: A survey on routing protocols for wireless sensor networks. Elsevier Ad Hoc Netw. **3**(3), 325–349 (2005)
4. Kavitha, M., Karthikeyan, K.: Comparison of data centric protocols for wsn and energy enhanced MSPIN(EEM-SPIN). Int. J. Eng. Res. Technol. (IJERT) **1**(9) (2012)
5. Heinzelman, W., Kulik, J., Balakrishnan, H.: Adaptive protocols for information dissemination in wireless sensor networks. In: Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom_99). Seattle, WA (1999)
6. Krishnamachari, B., Estrin, D., Wicker, S.: Modelling data centric routing in wireless sensor networks. In: Proceedings of IEEE INFOCOM, vol. 2, pp. 39–44 (2002)
7. Farivar, R., Fazeli, M., Miremadi, S.G.: Directed flooding: a fault-tolerant routing protocol for wireless sensor networks. In: Systems Communications Proceedings, pp. 395–399 (2005)
8. Zhang, Y., Fromherz, M.: Constrained flooding: a robust and efficient routing framework for wireless sensor networks. In: 20th International Conference on Advanced Information Networking and Applications, vol. 1, pp. 18–20 (2006)
9. Norouzi, A., Babamir, F.S., Zaim, A.H.: A novel energy efficient routing protocol in wireless sensor networks. J. Wirel. Sens. Netw. **3**, 341–350 (2011)
10. Yen, W., Chen, C.-W., Yang, C.-H.: Single gossiping with directional flooding routing protocol in wireless sensor networks. In: IEEE Conference on Industrial Electronics and Applications, pp. 1604–1609. Singapore, 3–5 June 2008
11. AlShawi, I.S., Yan, L., Pan, W., Luo, B.: A Fuzzy-Gossip routing protocol for an energy efficient wireless sensor networks. J. Sens. IEEE, pp. 1–4, 28–31 Oct 2012
12. Khanna, G., Bagchi, S., Wu, Y.-S.: Fault tolerant energy aware data dissemination protocol in sensor networks. In: International Conference on Dependable Systems and Networks, pp.795–804, June 28 2004, July 1 2004

Dynamic Satellite Network Routing Algorithm Based on Link Scheduling

Bingbing Zhang, Siqi Li, Wenjing Kang, Bo Li, Ruofei Ma, and Gongliang Liu(✉)

School of Information and Electrical Engineering, Harbin Institute of Technology, Weihai 264209, China
liugl@hit.edu.cn

Abstract. Inter-satellite link (ISL) in dynamic satellite network, due to its characteristics like wide coverage area, large population of users, and the motion of satellites in orbit planes, is easy to be interfered with. In order to solve above problems, this paper proposed a dynamic satellite network routing algorithm based on the link scheduling model. First, the geometric characteristics of ISL were analyzed, and the distance, elevation angle, azimuth of ISLs were calculated respectively based on spatial characteristics of constellation. Then we established a time-slot scheduling model of ISL based on TDMA system and proposed a routing algorithm of ISL based on hop threshold with a minimum transmission time. The final simulation results indicated that this routing algorithm can not only reduce the frequency of link switching in the path of information transmission effectively, but also complete most of the information transmission within half a superframe.

Keywords: Dynamic satellite network · Inter-satellite links · Link scheduling · Routing algorithm

1 Introduction

Satellite networking, which can greatly enhance the transmission performance of dynamic satellite networks, has become an important part of the global mobile communication system and is going to make global coverage come into reality. In recent years, with the increasing demand for space broadband communications, the study of key technologies in the new generation of broadband satellite communication networks with ISL and onboard processing functions, especially the study of routing algorithms have received widespread attention [1, 2].

Inter-Satellite Link (ISL) ISL is a link in the satellite network through which the exchange of information is completed via radio. It can achieve autonomous navigation by ranging and communication and improve the reliability of the system through cooperation with the ground segment [3, 4].

How to chose a routing between satellites is a crucial issue when using ISL for satellite networking. As for the satellite network, due to the constant movement of satellite nodes, the changes of network topology are too frequent to use ground routing

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 939–948, 2019.
https://doi.org/10.1007/978-981-13-6264-4_111

algorithm directly. Instead, a routing algorithm in accordance with the characteristics of satellite networks need to be designed.

At present, domestic and foreign scientific researchers have proposed a variety of satellite network routing algorithms for the characteristics of satellite networks. Chang et al. [5] proposed the Finite-state Automation (FSA) routing algorithm based on virtual topology. On the other hand, Ekici et al. [6] designed the datagram routing algorithm (DRA) for LEO satellite network.

This paper is organized as follows. In Sect. 2, we analyze the characteristics of topology of the MEO/GEO double-layer satellite constellation. A time-slot allocation model for ISL based on TDMA system is established in Sect. 3, and a routing algorithm which sets a threshold of the information transmission hop and selects the routing with minimum transmission time is designed in Sect. 4. In Sect. 5, stimulation is conducted to validate the performance of this routing algorithm. Conclusion is provided in Sect. 6.

2 Constellation Configuration and ISL Parameter Analysis

2.1 Constellation Configuration

This paper used GPS satellite constellation whose structure is 24/6/1 Walker constellation. 24 MEO satellites are distributed in six orbital planes with 55° of orbit inclination, 20200 km in orbital height and an orbital period of 11 h 58 min. In order to facilitate the analysis of data, 24 MEO satellites are named MEO11–MEO64 respectively. The high part of the number indicates the orbital plan number, and the low part of the number indicates the number of the satellite on the orbital plane. We also added three GEO satellites to the GPS satellite constellation to optimize the performance of satellite network. The three GEO satellites named GEO1, GEO2, and GEO3 are evenly distributed in the equatorial plane and the angle between each of them is 120°. Besides, a sensor is placed on each satellite to limit its visible range.

2.2 ISL Parameter Analysis

As the position of the satellites constantly changes, the geometric parameters of the ISL, such as the direction and distance, will also change. These time-varying parameters can have an impact on the allocation and routing algorithm of ISL.

The directivity of ISL is characterized by the following two parameters: elevation angle and azimuth between satellites. The definition of elevation angle and azimuth is shown in Fig. 1, in which satellites S_1, S_2 are non-geostationary satellites and they are not in the same orbital plane. Circle S_1 indicates the tangent plane when the satellite S_1 runs to the current position. The tangent plane intersects the plane OS_1S_2 (O is the center of the earth), and the straight line S_1P is the intersection line of them. Observed from the satellite S_1, angle $\angle S_1S_2P$ is the elevation angle. If it is specified that the azimuth of due north is zero degree, then the azimuth between satellite S_1 and satellite S_2 is greater than 180°.

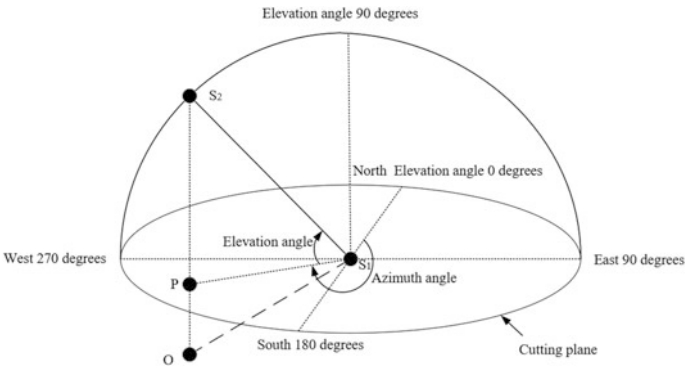


Fig. 1. The elevation angle and azimuth between Satellite S_1 and Satellite S_2

By calculating the specific values of the elevation angle and azimuth of the satellites, the directivity of ISL can be estimated. The calculation method is as follows:

Elevation angle ϕ_r:

$$\begin{aligned}\phi_r = -\frac{1}{2}\arccos\{&[\cos^2(\Delta\Omega/2) - \cos^2\beta\sin^2(\Delta\Omega/2)]\cos(\gamma_1 - \gamma_2) \\ &+ \cos\beta\sin\Delta\Omega\sin(\gamma_1 - \gamma_2) - \sin^2\beta\sin^2(\Delta\Omega/2)\cos(\gamma_1 + \gamma_2 + 2x)\end{aligned} \tag{1}$$

In formula (1), $\Delta\Omega$ represents the longitude difference between the ascending nodes of the orbits where the satellite S_1 and the satellite S_2 are located. $\gamma_1\gamma_2$ represents the phase difference between the two satellites at the initial time, and x means the phase of the satellite at the time t from its initial position.

Azimuth from satellite S_1 to satellite S_2 $(i = 1, j = 2)$:

$$\begin{aligned}\tan\psi_{ij} = &[\sin\beta\sin\Delta\lambda\cos(\gamma_j + x) - \sin 2\beta\sin^2(\Delta\lambda/2)\sin(\gamma_j + x)]/ \\ &[\sin^2\beta\sin^2(\Delta\lambda/2)\sin(\gamma_i + \gamma_j + 2x) + 2\cos\beta\cos(\Delta\lambda/2)\sin(\Delta\lambda/2)\times \\ &\cos(\gamma_i - \gamma_j) + \cos^2(\Delta\lambda/2) - \cos^2\beta\sin^2(\Delta\lambda/2)\sin(\gamma_i - \gamma_j)]\end{aligned} \tag{2}$$

According to the spherical triangle formula, the formula for the distance d between S_1 and S_2 on the sphere is

$$d = 2r|\sin(\phi_r)| \tag{3}$$

3 Dynamic Satellite Network Link Scheduling Strategy

3.1 ISL Visibility Processing

Visibility between two satellites is a physical prerequisite for establishing ISL. As satellites are in a state of constant motion, the relative positions of them will change with time, which leads to dynamic changes of the visibility between satellites.

Based on TDMA system, this paper designs the ISL scheduling scheme and establishes the ISL time-slot scheduling model. The orbital period of satellite is divided into a number of equally spaced superframes, then each superframe is further divided into a number of equally long time slots.

Definition 1 The inter-satellite visibility in the k-th superframe is defined as

$$v_{ij} = \begin{cases} 1, \textit{satellite } S_i \textit{ and } S_j \text{ are continuously visible at the k} - \text{th superframe} \\ 0, \textit{satellite } S_i \textit{ and } S_j \text{ are intermittently visible at the k} - \text{th superframe} \end{cases} \tag{4}$$

$$V_k = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1M} \\ v_{21} & v_{22} & \cdots & v_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ v_{N1} & v_{N2} & \cdots & v_{NM} \end{bmatrix}_{N \times M} \tag{5}$$

V_k is the inter-satellite visibility matrix of the k-th superframe. N is the total number of satellites in the satellite network, and M is the number of time slots per superframe.

Definition 2 Internal satellite and External satellite: Ground stations are set up in the satellite scene for transmitting signals to and receiving signals from satellites. Then, satellites visible to the ground station are defined as the internal satellites, and satellites which are invisible to the ground station are the external satellites.

Definition 3 Static ISL and Dynamic ISL: ISLs between satellites which are visible throughout the satellite operation cycle are defined as static ISLs, and ISLs between satellites which are intermittently visible during the operation cycle are defined as dynamic ISLs.

Static ISL has the advantages of wide distribution, large elevation angle, and high link budget, so its performance is better than that of dynamic ISL.

3.2 ISL Time-Slot Scheduling Model

This ISL time-slot scheduling model used is based on TDMA system. In each time slot: (1) There are mutually uninterrupted links between the satellites; (2) Each satellite can only have one link of them [7].

First, in order to facilitate data analysis and collation, satellites MEO11–MEO64 are numbered from1 to 24, and satellites GEO1–GEO3 are numbered from 25 to 27. The links of satellite S_i are classified, of which the static ISLs are put into set Q_i and the dynamic ISLs are put into set D_i.

Second, for the k-th superframe, the ISL time- slot matrix P_k is established. Because the performance of the static ISL is better than that of the dynamic ISL, the static ISL is preferentially placed into the time-slot matrix. The main function of the ISL is to relay information between invisible internal satellites and external satellites. Therefore, the smaller the probability that they exist at the same time, the better.

According to the order of the probability that the satellites are present in the territory at the same time from small to large, the satellites in set Q_i are sequentially placed in the i-th row of the time-slot matrix P_k. The satellites in set D_i are then sorted in the same way as described above and are placed in the subsequent time slot of the i-th row of slot matrix P_k. The slot matrix P_k is represented as

$$P_k = \begin{bmatrix} a & d & \cdots & f \\ c & 0 & \cdots & e \\ \vdots & \vdots & \ddots & \vdots \\ 0 & r & \cdots & g \end{bmatrix} = \left[\lambda_{ij}\right]_{m\times n} \tag{6}$$

In formula (6), the element λ_{ij} in matrix P_k represents the satellite number pointed to by satellite i in the j-th time slot of the k-th superframe. If λ_{ij} is zero, it means that satellite i did not establish an ISL in the k-th time-slot of the k-th superframe. m represents the number of satellites in the constellation, and n represents the number of time slots in a superframe.

Then, based on the inter-satellite visibility matrix V_k in this superframe, the inter-satellite links between the invisible satellites in the k-th superframe in the time-slot matrix P_k are deleted, and the corresponding position elements are set to zero. Finally, the GEO satellites visible with the MEO satellites in the k-th superframe are placed in the position of the zero elements in the time-slot matrix P_k according to certain rules. At this point, the ISL time-slot allocation matrix of the k-th superframe is obtained.

4 ISL Routing Algorithm

4.1 Information Transmission Time Analysis

At present, various researches are basically based on virtual topology control strategies to study routing algorithms. However, the ISL scheduling model established in this paper is based on the TDMA system, and inter-satellite links are arranged in fixed time slots according to the rules of the scheduling model. When information is routed according to the routing algorithm, there is a double limitation of space and time. Therefore, the information transmission time is no longer only related to the distance between satellite nodes. For the ISL time-slot scheduling model of this paper, the time from starting looking for the dissemination path to arriving the destination satellite node includes the following three parts:

(1) The delay t_w when the source satellite node waits to send information;
(2) The delay t_a when the information dissemination in free space;

(3) The number of time slots t_r that the information spent on the transmission path

Three delays are not the simple superposition relationship. The second delay t_a is much less than the third delay t_r, and t_a and t_r are inclusion relations. Therefore, the total time of information transmission is

$$t_{delay} = t_w + t_r \tag{7}$$

4.2 ISL Routing Algorithm Steps

For ISL, link stability is a very important but difficult problem to solve. The routing algorithm studied in this paper can effectively solve the contradiction between link switching times and information transmission time. That is, within a limited number of hops c, the minimum transmission time is found, and an inter-frame transmission mechanism is added to solve the packet loss caused by the link switching between each frame. The specific algorithm process is as follows:

step 1 For the satellite scene studied in this paper, the visibility relationship between satellites is analyzed and the number of time slots and the length of time slots in a frame are set reasonably. Based on the analysis results, the ISL time-slot scheduling model is established, and the time-slot matrix P for each frame in the constellation operation cycle is obtained.
step 2 According to the slot matrix P, the simulation finds a reasonable limited hop count c so that the hop count value not only satisfies the proper number of link switching, but also enables the information to find the path from the source satellite node to the destination satellite node.
step 3 The source satellite node S_s numbered S generates information at time t and needs to transmit the information to the destination satellite node S_d numbered d. Calculate the superframe number k and time-slot number q where time t is located. First, look for all one hops in the k-th superframe to reach the destination satellite node's path set L_1, then look for all two hops to reach the destination satellite node's path set L_2……until you find all c hops to reach the destination satellite node's path set L_c. Calculate the total transmission time of each path in all sets $t_{delay1}, t_{delay2}, \ldots\ldots, t_{delayn}$. The corresponding paths are $l_{t_{delay1}}, l_{t_{delay2}}, \ldots\ldots, l_{t_{delayn}}$. Select the path with the shortest total transmission time as the final information propagation path, namely:

$$L = \left\langle l_{t_{delay1}}, l_{t_{delay2}}, \ldots\ldots, l_{t_{delayn}} \right\rangle \tag{8}$$

If no transmission path is found in the k-th superframe, go to the fourth step.

step 4 If no dissemination path is found in k-th superframe, the inter-frame transmission mechanism is started. The information is generated in time slot numbered q and the first nonzero element after time slot q in the s-th row of the time-slot matrix P_k is sought—satellite S_{s1}, which is considered as the first hop of the path and is treated as a temporary source satellite node. Then, the shortest transmission time

path from the temporary source satellite node S_{s1} to the destination satellite node S_d is found in the superframe numbered $k + 1$ according to the method of the third step.

4.3 Multiple Service Information Queuing Processing Mechanism on the Satellite

When multiple service information are transmitted in the satellite network, it is highly likely that multiple service information arrive at a satellite node at the same time. If the routing algorithm does not include the multiple service information queuing processing mechanism, channel congestion is likely to occur, resulting in increased transmission delay or even packet loss. For this situation, the following mechanism is used for processing:

1. If a plurality of service information simultaneously start from a certain source satellite node, the transmission paths of the plurality of service information are sequentially calculated according to the order of arrival at the source satellite node. Then according to the path allocated by the routing algorithm, the service information with less waiting time slots is transmitted first. If multiple service information need to use the same time slot, they are transmitted in order according to the sequence of arrival at the satellite.
2. If multiple service information arrive at a certain intermediate satellite node at the same time according to paths allocated by the routing algorithm, the information that can be sent in the time slot closest to the current time is transmitted first according to the original route calculated by the routing algorithm. If multiple service information need to use the same time slot, they are transmitted in order according to the sequence of arrival at the satellite. If the queuing delay of a service information exceeds the time threshold set in the algorithm, the intermediate satellite node recalculates the transmission path to avoid packet loss.

5 Simulation and Analysis

This paper makes use of STK and MATLAB software to verify the designed routing algorithm. Based on the stability of ISL, the time and number of hops of the information are briefly evaluated.

The service flow network scenario set in the simulation is used to verify the routing algorithm's ability to unicast service flow. Since the Walker constellation established in this paper has symmetry, the satellites MEO11 and MEO13 on the same orbit plane, the satellites MEO13 and MEO63 on the different orbit planes are selected respectively as the source satellite node and the destination satellite node of the service information flow in this scenario. During the satellite constellation operation cycle, packets are evenly sent from the source satellite node at a packet sending rate of 20 packets/min. The calculation time of the routing algorithm is negligible. The ISL routing algorithm designed in this paper is used to calculate the transmission path of the packet. If the multi-service information arrive at an intermediate satellite node at the same time, the

onboard service information queuing processing mechanism is started. Simulation shows the average transmission time probability histogram and the average transmission hop probability histogram of the service information flow in the satellite constellation operation cycle, as shown in Figs. 2 and 3.

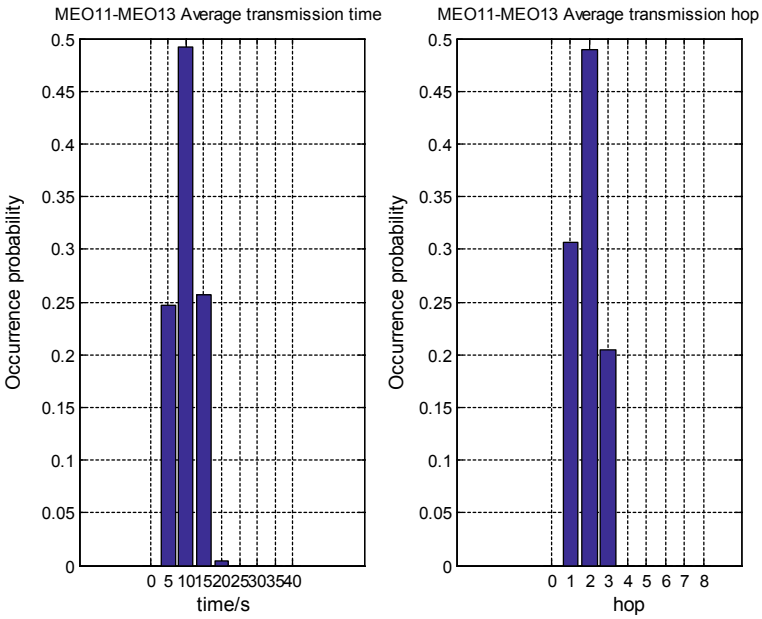


Fig. 2. Average transmission time probability histogram and the average transmission hop probability histogram of the service information flow from MEO11 to MEO13

From Figs. 2 and 3, the average transmission time of the service information flow between the satellites in the same orbit is 5–20 s, and the proportion of 10–15 s is very large. The average transmission time of the service information flow between the satellites in the different orbits is 5–25 s. The proportion of the same 10–15 s accounts for a large proportion. Regardless of whether the information is transmitted between satellites in the same orbit or between satellites in the different orbit, their average transmission hops are between 1 and 3 hops. The number of inter-satellite link switching is measured by the number of hops of information. This shows that the ISL routing algorithm studied in this paper ensures that most information can be transmitted from the source satellite node to the destination satellite node through a small number of link switching. And because of the inter-frame transmission mechanism, the shortest transmission time path of information is always found from the source satellite node to the destination satellite node according to routing algorithm, even if it is generated in the last time-slot of each superframe, which greatly improves the stability of ISLs.

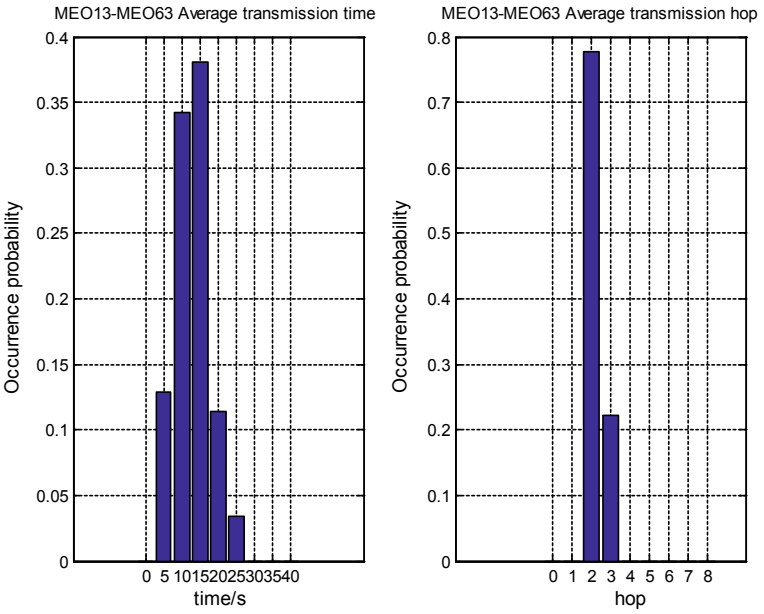


Fig. 3. Average transmission time probability histogram and the average transmission hop probability histogram of the service information flow from MEO13 to MEO63

6 Conclusion

In this paper, a time-slot scheduling model for ISL is established based on TDMA system, and a new dynamic satellite network routing algorithm is designed for this model. The routing algorithm sets a threshold of information propagation hops, and then finds the path with shortest transmission time, which enables most of the information generated at any time during the satellite's orbital period to be transmitted in a limited number of hops within half a superframe. The inter-frame transmission mechanism is also added to this routing algorithm to reduce the impact of link switching on information transmission. After simulation analysis, this algorithm is proved to greatly improve the stability of information transmission in dynamic satellite networks. This dynamic satellite network routing algorithm designed in this paper is only a preliminary result. Further study and optimization of the routing algorithm will be conducted in the near future. For instance, in order to enhance the performance of information transmission, services with different priories will adopt different rules to choose routing.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (No.61371100, No.61501139, No. 61401118).

References

1. Wan, P., Ye, J., Song, S.: Study on the telecommunication technology based on the distributed satellite constellation networks. In: IEEE Aerospace Conference, pp. 1–8. IEEE (2014)
2. Hong, S.G., Su, C.J.: ASAP: fast, controllable, and deployable multiple networking system for satellite networks. In: IEEE Global Communications Conference, pp. 1–7. IEEE (2014)
3. Holmes, J.K., Raghavan, S.: A summary of the new GPS IIR-M and IIF modernization signals. In: Vehicular Technology Conference, Vtc2004-Fall, vol. 6, pp. 4116–4126. IEEE Xplore (2005)
4. Geng, L., WU, S.Q.: Constellation design and establishment of inter-satellite links. Chin. Space Sci. Technol. **20**(6), 62–67 (2000)
5. Chang, H.S., Kim, B.W., Lee, C.G., et al.: FSA-based link assignment and routing in low earth orbit satellite networks. IEEE Trans. Veh. Technol. **47**(3), 1037–1048 (1998)
6. Ekici, E., Akyildiz, I.F., Bender, M.D.: A distributed routing algorithm for datagram traffic in LEO satellite networks. IEEE/ACM Trans. Netw. **9**(2), 137–147 (2001)
7. Wu, G.Y., Chen, J.Y., Guo, X.Y., et al.: Distribution design and simulation of time-slot allocation of inter-satellites links based on TDMA. Comput. Meas. Control **22**(12), 4087–4090 (2014)

Research on Intelligent Anti-jamming Communication with Transform Domain Communication System

Meng Liu, Jingning Wang(✉), and Lizhe Liu

The 54th Research Institute of CETC, No.589 West Zhongshan Road, Shijiazhuang, China
cike009@qq.com, wjnhit@126.com, liu_lizhe@sina.com

Abstract. In order to solve the interference problem in a complex electromagnetic environment, NC-OFDM and TDCS are efficient transmit waveform. The spectral efficiency and the applicable environment of TDCS are different from NC-OFDM. The simulation results show that NC-OFDM can obtain the higher spectral efficiency and the required SINR is higher, and TDCS can work well in lower SINR environment with different spreading factor. When there is little spectrum hole, spread spectrum is a better choice. There is no threshold of some parameter which can make a decision. An artificial intelligence algorithm is proposed to select the optimal one from several waveforms. The input parameters of the algorithm include SINR, interference bandwidth, spectrum holes, etc.

Keywords: Intelligent anti-jamming communication · Transform domain communication system · Artificial intelligence algorithm

1 Introduction

In a complex electromagnetic environment, cognitive radio technology is an effective means to resist interference. Both Non-Contiguous OFDM (NC-OFDM) and Transform Domain Communication System (TDCS) are waveforms suitable for discrete spectrum situations of cognitive radio scenarios, they can obtain real-time spectrum and interference distribution through spectrum sensing. If it is difficult to find the available spectrum holes, spread spectrum is a suitable anti-interference waveform [1].

From Fig. 1, it can be seen the energy distribution of NC-OFDM and TDCS is in the frequency band from which interference is removed. First, the spectrum sensing module is required to provide accurate spectrum information, and the unavailable frequency band (occupied or interfered by other systems) is removed to achieve spectral hole access to the spectrum hole.

Because the interference signal is often time-varying, the available spectrum is often discrete and irregular [2]. This requires the study of smart anti-jamming technology. Intelligent anti-jamming technology is the application of intelligent communication in the field of anti-jamming communication. According to the electromagnetic interference environment, it intelligently generates the best anti-jamming method,

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 949–955, 2019.
https://doi.org/10.1007/978-981-13-6264-4_112

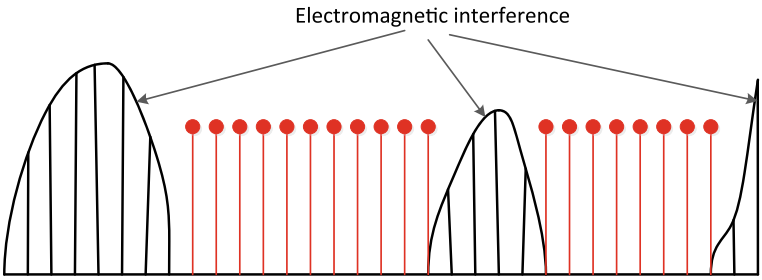


Fig. 1. NC-OFDM/TDCS energy distribution

improves the system's anti-jamming capability and spectrum utilization, and achieves efficient and reliable anti-jamming communication. Since the result of spectrum sensing is a random sequence, it is difficult to determine the corresponding waveform with several parameters. Therefore, this paper explores the use of artificial intelligence algorithms such as neural networks to select the best waveform.

2 Communication Waveform and Artificial Intelligence Algorithm

2.1 NC-OFDM Waveform

The basic principle of NC-OFDM is shown in Fig. 2, which includes spectrum sensing, coding and decoding, modulation and demodulation, OFDM modulation and demodulation, SNR estimation, channel estimation, and other modules. Compared with the traditional OFDM technology, the basic structure of NC-OFDM is that there is more dynamic spectrum sensing at the transceiver end to eliminate the unavailable subcarriers [3].

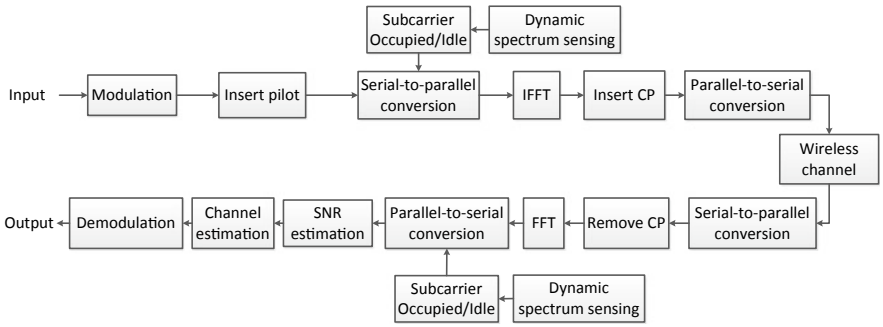


Fig. 2. The basic principle of NC-OFDM

2.2 TDCS Waveform

TDCS is a cognitive radio technology. Combined with the spread spectrum approach, the useful signal is spread over the entire available frequency band with a low power spectral density.

The basic principle of TDCS is shown in Fig. 3. First, the environmental spectrum is scanned, and the target frequency band is quantized into N sub-carriers so as to facilitate hardware implementation. Then, by comparing the spectral amplitude of the scan result with a preset threshold and marking the availability of the k-th sub-carrier as A_k. If the spectrum amplitude at the available frequency points is less than the threshold value, note $A_k = 1$, It is considered that the interference within this frequency range can be used directly; otherwise, $A_k = 0$, it is considered that there is a large amount of interference here, or there is an authorized user conducting communication, where the spectrum is not available. In the above manner, the transceiver obtains all the sub-carrier utility sequences in the TDCS system, denoted by $A = \{A_0, A_1, A_2, \ldots, A_{N-1}\}$. This sequence determines the spectral shape of the TDCS signal.

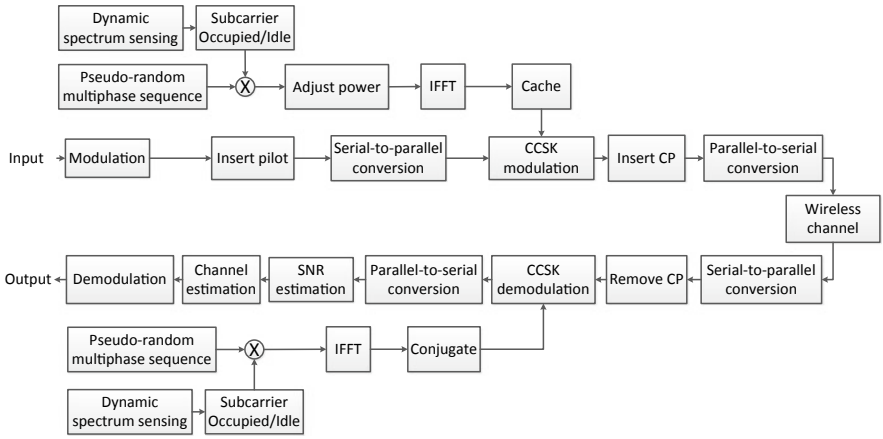


Fig. 3. The basic principle of TDCS

The transmitter first generates a set of N-point long pseudorandom multi-phase sequences through a random phase mapper, $P = \left\{e^{jm_0}, e^{jm_1}, \ldots e^{jm_{N-1}}\right\}$. The phase mapping method can be generated using a Linear Feedback Shift Register (LFSR). Then, the obtained pseudorandom phase sequence is multiplied by the above-mentioned spectrum vector A point by point to obtain the frequency domain coding sequence of the TDCS, i.e.,

$$B = [B_0, B_1, \ldots B_k, \ldots B_{N-1}] = \lambda \bullet A \otimes P \tag{1}$$

$$B_k = \lambda A_k e^{jm_k}, k \in \{0, 1, \ldots, N-1\} \tag{2}$$

Among them, $\otimes$ represents the dot multiplication operation, $\lambda = \sqrt{\varepsilon_s/N_A}$ is the power normalization adjustment factor, ε_s is the energy needed to transmit one symbol, and NA is the number of all $A_k = 1$. The frequency domain sequence s makes the TDCS signal have the following two characteristics: (1) The frequency domain amplitude only has a valid value at the position where the frequency spectrum decision is available, and the remaining positions are zero, thereby effectively avoiding the interference; (2) The multi-phase information makes The signal waveform of TDCS has pseudorandom noise-like correlation characteristics, which is the anti-intercept feature of the TDCS and the technical foundation of multiple access. By performing an inverse fast Fourier transform on B, its time-domain expression is obtained

$$b = \{b_0, b_1, \ldots b_n, \ldots b_{N-1}\} = \mathrm{IFFT}\{B\} \tag{3}$$

$$bn = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} B_k e^{\frac{j2\pi kn}{N}} = \frac{\lambda}{\sqrt{N}} \sum_{k=0}^{N-1} A_k e^{jm_k} e^{\frac{j2\pi kn}{N}} \tag{4}$$

The b in the above equation is also called the basic modulation waveform and is buffered for the subsequent modulation operation to continue to use.

A typical modulation method in TDCS is a circular shift keying, which uses a basic modulation waveform as a basis function to realize modulation transmission of effective data bits. Corresponding receivers use a similar RAKE receiver for signal demodulation.

2.3 Spread Spectrum Waveform

The information input by the spread spectrum communication transmitting terminal is modulated by information to form a digital signal, and then the digital signal is modulated by the spreading code sequence generated by the spreading code generator to broaden the spectrum of the signal. The signal after widening modulates the carrier frequency again.

The spreading factor is an important parameter in spread spectrum communication. It is the number of chips each symbol is converted into. It is also the ratio of the post-spreading chip rate to the pre-spreading signal rate, which directly reflects the spreading gain.

2.4 Artificial Intelligence Algorithm

Neural network algorithm is used to realize intelligent anti-jamming decision algorithm (Fig. 4). The input parameters include spectrum sensing information, communication demand parameters, other constraints, and some knowledge bases composed of measured or simulated data. The spectrum sensing information is the energy at each frequency point of the entire frequency band obtained. The interference bandwidth, interference intensity, signal-to-noise ratio of the noninterference area, and other parameters can be obtained. Communication demand parameters include expected speed and reliability; other constraints include hardware complexity and other

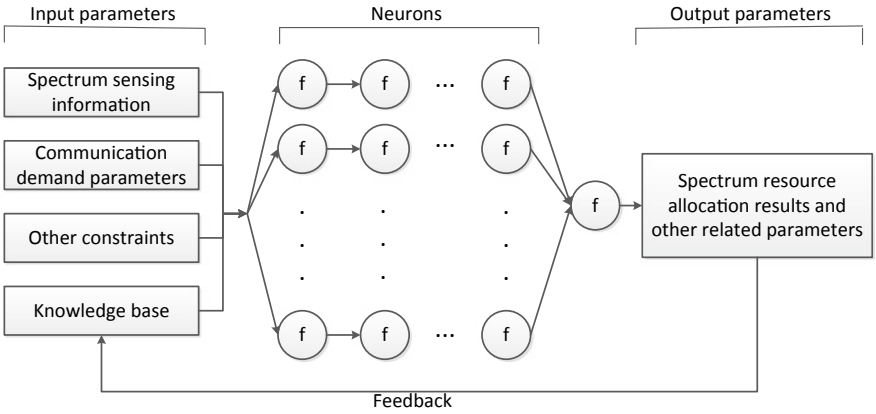


Fig. 4. The basic principle of artificial intelligence algorithm

parameters; knowledge base includes known training data and original parameters obtained from actual measurement or simulation. The output parameters obtained after passing through a plurality of neuron function models include waveform selection results and corresponding parameters (such as modulation method of OFDM, basis function of TDCS, spreading factor of spread spectrum, etc.) (Fig. 4) [4].

3 Simulation Results

3.1 Parameter Setting

The setting of this simulation parameter is

Transmission bandwidth: 20 MHz
Number of sub-carrier: 256
Interference to signal ratio: 10 dB
OFDM: FFT points are 256 and cyclic prefix length is 64
TDCS: the basic modulation waveform frequency is 256, FFT points is 256, the cyclic prefix length is 64
Spread Spectrum factor: 256
BER requirement: 10^{-3} (regardless of encoding)

3.2 Simulation Comparison

Figures 5 and 6 are simulation results in different interference. The upper left of Figs. 5 and 6 are simulation results without interference. Simulation results show that when the SNR is above 10 dB can use OFDM waveforms for high-speed data transmission, if not we can only use the TDCS or spread spectrum.

The upper right of Figs. 5 and 6 are simulation results when there is interference in 128 sub-carriers. Simulation results show that OFDM waveform cannot meet the BER

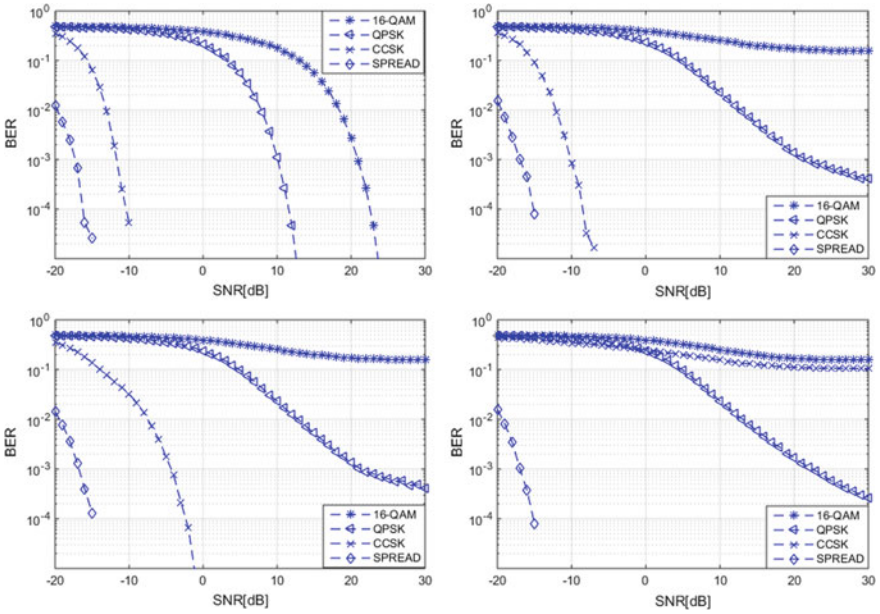


Fig. 5. BER simulation of different interference

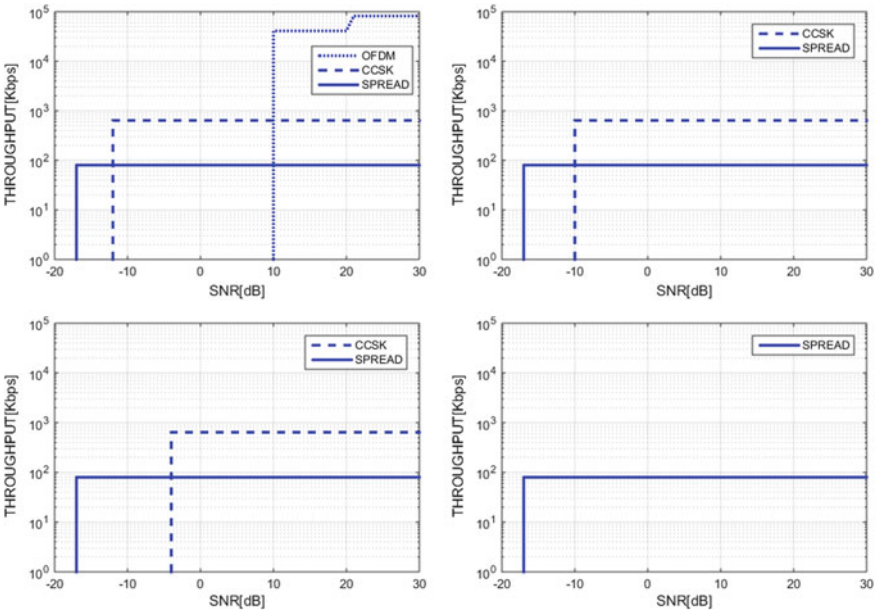


Fig. 6. Throughput simulation of different interference

requirements, TDCS waveform can meet the maximum rate requirements when SNR is above −10 dB, and spread spectrum waveform can be used in lower SNR.

The lower left of Figs. 5 and 6 are simulation results when there is interference in 192 sub-carriers. The required SNR is above −4 dB with TDCS.

The lower right of Figs. 5 and 6 are simulation results when there is interference in 250 sub-carriers. OFDM. TDCS cannot meet BER requirement, and only spread spectrum waveform can be used.

Since there is no definite correspondence between the results of spectrum sensing and the results of waveform selection, it is necessary to extract the characteristic values of the spectrum sensing results, and then use an artificial intelligence algorithm to judge the selected waveform and parameters in real time.

4 Conclusion

Both NC-OFDM and TDCS can eliminate the interfered and occupied sub-carriers through the spectrum sensing module. NC-OFDM can achieve high-speed transmission in the case of small narrowband interference. The TDCS can achieve reliable transmission of data in the case of large narrowband interference. Spread spectrum can achieve the reliability of data transmission without spectrum cavitation. The electromagnetic environment is becoming more and more complex and changes rapidly. It is difficult to solve the problem of maximizing the communication effect under complex interference situations by simply relying on interference waveforms such as TDCS, and the combination of multiple waveforms through intelligent algorithms is an important means of future communication anti-jamming technology.

Acknowledgements. This work is supported by Project funded by China Postdoctoral Science Foundation under Grant No.2016M601286, the fund of Science and Technology on Communication Networks Laboratory. Moreover, this work is also supported by the fund of National Key Laboratory of Science and Technology on Communications under Grant number EX17641Z080.

References

1. Budiarjo, I., Nikookar, H., Ligthart, L.P.: Cognitive radio modulation techniques. IEEE Signal Process. Mag. **25**, 24–34 (2008)
2. Li, S., Cheng, Y., Dong, B., et al.: Research on intelligent anti-jam communication techniques. Radio Commun. Technol. **38**(1), 1–4 (2012)
3. Chakravarthy, V., Nunez, A.S., Stephens, J.P., et al.: TDCS, OFDM, and MC-CDMA: a brief tutorial. IEEE Commun. Mag. **43**, S11–S16 (2005)
4. Tanweer, M.R., Auditya, R., Suresh, S., et al.: Directionally driven self-regulating particle swarm optimization algorithm. Swarm Evol. Comput. 24–30 (2016)

Bandwidth Allocation and Performance Analysis for Ad Hoc Based UWA Networks

Liu Jingjing[1(✉)], Li Chunguo[2], Song Kang[3], and Liu Chuanyang[1]

[1] Chizhou University, Chizhou, Anhui, China
59916185@qq.com, 280176079@qq.com
[2] Southeast University, Nanjing, China
[3] Qiingdao University, Qingdao, China

Abstract. The direct point-to-point long distance underwater acoustic communication suffers significant attenuation due to the complex underwater environment. Underwater acoustic wireless self-organized network is developed to improve the transmission quality by introducing forwarding nodes. In this paper, the bandwidth allocation and performance analysis for flat distributed network structure are studied, where network MAC constraints are considered. Linear programming algorithm is deployed to obtain the optimal bandwidth allocation. Through simulation, the SNR and BER performance gain compared with direct point-to-point communication are validated.

Keywords: Underwater acoustic ad hoc · MAC constraint · Linear programming · Signal-to-noise ratio · Bit error rate

More and more discussions about underwater acoustic communication. The medium for underwater acoustic communication is sound waves [1–3], especially long-distance communications underwater is difficult. The noise, caused by fish, ships, and waves in the ocean, makes the sound field there extremely chaotic, blurring received signals [4]. Since the current underwater communication mostly adopts point-to-point communication, it is difficult to meet the need of communication on the sea and inside the sea. So there is an urgent need to develop a new underwater communication network. From underwater ad hoc network, this paper discusses and analyzes the influence of bandwidth allocation algorithm of linear programming on the performance of underwater acoustic communication.

1 Overview of Ad Hoc Based UWA Networks

The underwater acoustic communication network can be flexibly used in different seabed and sea surface scenarios, and is widely used in marine environment observation and so on. It can exchange information at multiple observation points in different space and position in water. Moreover, the underwater acoustic channel has a variety of

Natural Science Foundation of Chizhou University (2016ZRZ006), Chizhou University Teaching and Research Project (RE18200030). National Natural Science Foundation of China (61671144), Anhui quality engineering project (2014zy077).

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 956–963, 2019.
https://doi.org/10.1007/978-981-13-6264-4_113

transmission states, and the operating environment of the ocean is relatively bad. It has a high demand for the reliability of the communication algorithm and equipment, and the networking of underwater acoustic communication has become a hot research topic at present [5–7]. The idea of ad hoc network originated from ALOHA network in the United states. Ad hoc network has the following characteristics: dynamic topology, self-organizing structure, distributed network, multi-hop routing.

Underwater acoustic communication transmits information mainly through water, its sound transmission medium, instead of air. The communication environment under water is more complex than in the air, acoustic wave spreads in water is difficult, and it spreads far in the air, in order to increase the acoustic communication distance can increase the communication nodes by forwarding the information, the information can be transferred more far, this is the ad hoc network for communication under water. There are two kinds of networking modes in ad hoc networks: hierarchical distributed architecture and planar distributed architecture, which are shown in Figs. 1, 2, respectively. The different characteristics of the two structures are suitable for different occasions, and the underwater communication occasions are more suitable for second structure, because the nodes in the second structure are freely joined, and the network build faster.

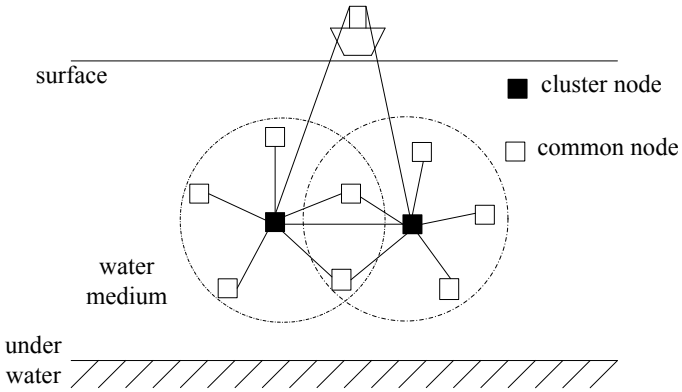


Fig. 1. Layered distributed network structure

Because underwater acoustic communication is different from electromagnetic wave communication, the routing protocol and bandwidth allocation algorithm of underwater acoustic ad hoc network are different from ad hoc network in air. This paper mainly analyzes the bandwidth allocation of underwater acoustic ad hoc network.

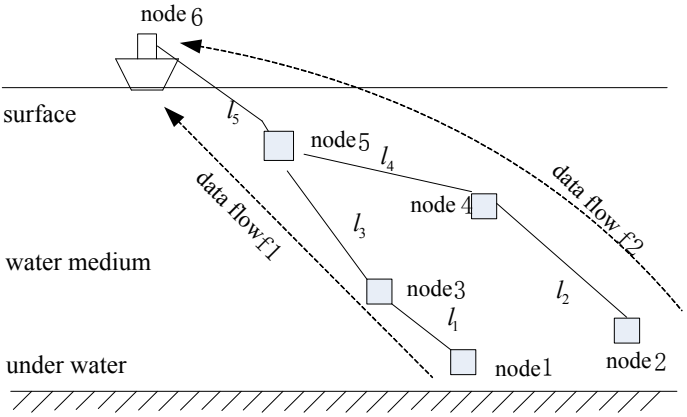


Fig. 2. Planar distributed network structure

2 MAC Constraints in Underwater Acoustic Communication

When the node communication in the multi-hop ad hoc network is communicated [8], the problem of congestion control [9] should be considered. In another word, the constraint of MAC should be considered.

MAC constraints [10] include link constraints and time slot constraints. Link constraint means that the sum of the rates of data streams passing through a link can not exceed the bandwidth capacity of the link. The time constraint is the communication and reception data of all communication nodes in the network at the same time. Take Fig. 2 as an example to discuss MAC constraints.

For each node $N_i(i = 1, 2, 3, 4, 5, 6)$, two data streams $f_k(k = 1, 2)$ are transmitted to the terminal node 6 on the water surface node 1 and node 2 respectively, and two data streams are transmitted through the intermediate node respectively. The channel capacity of each link is $C_j(j = 1, 2, 3, 4, 5)$.

Which nodes are passed through each node and are represented by matrix A. The row number of the matrix represents the node number. The column number represents the link number. A link passes through a node and is represented by 1. It does not go through 0. For example $A_{ij} = 1$ means a link l_j passes through a node N_i, and if it does not pass, the value is equal to 0.

The actual use rate of node time slot is expressed in $\varepsilon_i \in [0, 1]$, that is, node utility factor, and the utility factor of each node is expressed by $G = (\varepsilon_1, \ldots \varepsilon_i)$.

The link that the data flow passes is represented by matrix B. The line number of the matrix represents the link number. The column number represents the ordinal number of the data stream. If a data stream flows through a link, the value of the corresponding position of the matrix is the reciprocal of the capacity of the channel, otherwise the value is zero, for example,

$B_{jk} = \frac{1}{C_j}$ indicates that the data flow $f_k(k = 1, 2)$ passes through the link l_j, and if it is zero, it is not passed through the link.

The transmission rate of the data flow f_k is expressed in $X = (x_1, \ldots x_k)$.On the time slot constraint equation for MAC is $ABX' \leq 1 - G'$.

Take the network of Fig. 2 as an example,

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} \frac{1}{C_1} & 0 \\ 0 & \frac{1}{C_2} \\ \frac{1}{C_3} & 0 \\ 0 & \frac{1}{C_4} \\ \frac{1}{C_5} & \frac{1}{C_5} \end{bmatrix} \tag{2}$$

$$X' = (x_1, x_2)', 1 - G' = (1 - \varepsilon_1, 1 - \varepsilon_2, \ldots 1 - \varepsilon_6)', ABX' \leq 1 - G'$$

The time slot constraint is equivalent Eq. (3)

$$\begin{bmatrix} \frac{x_1}{C_1} \\ \frac{x_2}{C_2} \\ \frac{x_1}{C_1} + \frac{x_1}{C_3} \\ \frac{x_2}{C_2} + \frac{x_2}{C_4} \\ \frac{x_1}{C_3} + \frac{x_1}{C_5} + \frac{x_2}{C_4} + \frac{x_2}{C_5} \\ \frac{x_1}{C_5} + \frac{x_2}{C_5} \end{bmatrix} \leq \begin{bmatrix} 1 - \varepsilon_1 \\ 1 - \varepsilon_2 \\ 1 - \varepsilon_3 \\ 1 - \varepsilon_4 \\ 1 - \varepsilon_5 \\ 1 - \varepsilon_6 \end{bmatrix} \tag{3}$$

The link constraint on MAC is that the sum of the rate of each data flow of a link should not excess the bandwidth capacity of the link. The expression is as follows:

$$\sum_k x_k \leq C_j \tag{4}$$

Take Fig. 2 as an example, the link constraint on MAC is:

$$\begin{cases} x_1 \leq C_1 \\ x_1 \leq C_3 \\ x_2 \leq C_2 \\ x_2 \leq C_4 \\ x_1 + x_2 \leq C_5 \end{cases} \tag{5}$$

3 Linear Programming Bandwidth Allocation Algorithm Based UWA Networks

The algorithm finds the minimum elements of all rate vectors in the feasibility set of bandwidth allocation at every step, and maximizes the minimum element by using the optimization theory of linear programming [11]. The minimum element is fixed, and the next step is to find the second smallest element of the maximum minimum fair rate vector in the feasibility set. In turn, each element $x(x_1, x_2, \ldots x_i, \ldots x_j, \ldots x_n)$ in the

bandwidth allocation vector x_i is determined. The set of feasibility here refers to a set of MAC constraints that are satisfied by the velocity of each data. The flow chart of the algorithm is as Fig. 3. In this algorithm, each vector is maximized as far as possible. That is the maximum bandwidth allocated to each data stream is guaranteed. Not only this algorithm guarantee the maximum throughput, but also the performance index can improve connectivity.

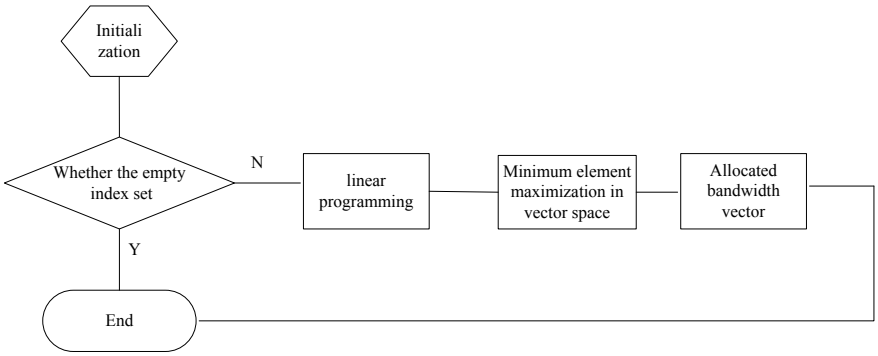


Fig. 3. Flow graph of linear programming algorithm

4 Performance Analysis

The application of ad hoc communication network mode based on MAC constraint and linear programming bandwidth allocation algorithm to underwater acoustic communication will improve the performance index of underwater acoustic communication, signal-to-noise ratio, the bit error rate, and reduce the outage probability of signal transmission.

(1) Signal-to-noise ratio

Signal-to-noise ratio (SNR) is a very important performance index in communication system. In the shallow sea multipath channel, the signal-to-noise ratio estimation algorithm is used to estimate the signal-to-noise ratio [12], which is very close to the actual signal-to-noise ratio. It is assumed that the received signal is A, and the signal-to-noise ratio is as follows:

$$SNR = \frac{h_{yy}(kT_S)}{J_0(2\pi f_m k)\cos(2\pi f_c k T_s) - h_{yy}(kT_s)}, k = 1, 2, 3 \cdots \tag{6}$$

$$h_{yy}(kT_S) = \frac{e^2 J_0(2\pi f_m k)\cos(2\pi f_c k T_s)}{e^2 + e_n^2}, k = 1, 2, 3 \cdots, \tag{7}$$

The e^2 represents the signal power, and the e_n^2 represents the noise power. $J_0(x)$ represents the first class zero order Bessel's function, and f_m represents the normalized Doppler shift.

If the signal frequency is 5 kHz, the simulation of average SNR between direct communication and ad hoc communication is for Fig. 4, the signal-to-noise ratio is increased when the communication nodes is increased, with the increase of the communication distance, the signal-to-noise is increased more obvious. Figure 5 is the simulation curve about SNR of different distances at the 5 km distances. The more intermediate nodes, the more SNR is increased.

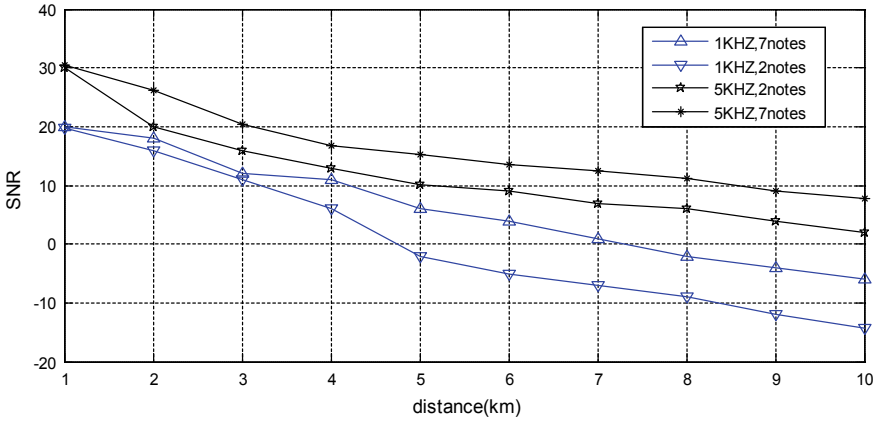


Fig. 4. SNR of different distances

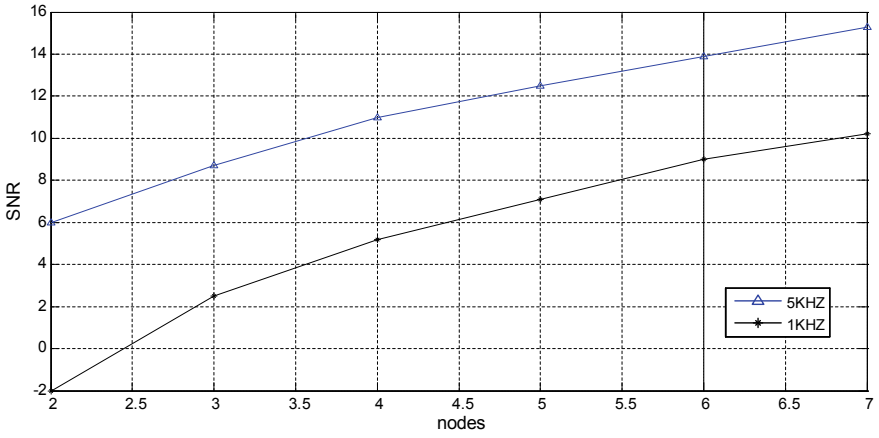


Fig. 5. SNR of different nodes

(2) error rate

The error probability in underwater acoustic communication has a great influence on the quality of communication, and it can be used to measure the reliability of the system. The error rate of underwater acoustic channel is related to the propagation characteristics of the channel. Using ad hoc network, it can improve the channel propagation characteristics between two communication nodes, and also play a role in improving the bit error rate.

If the coherent FSK is used in the shallow sea channel, the bit error rate is

$$p_e = \frac{1}{2}\left[1 - \frac{1}{\sqrt{1+\frac{4}{W}}}\right]$$

If PSK is used, the bit error rate is $p_e = \frac{1}{2}\left[1 - \frac{1}{\sqrt{1+\frac{2}{W}}}\right]$, the $W = e^2/e_n^2$, FSK simulation is used here, as shown in Fig. 6, it can be seen that adding intermediate nodes can effectively reduce the bit error rate and improve the communication effect.

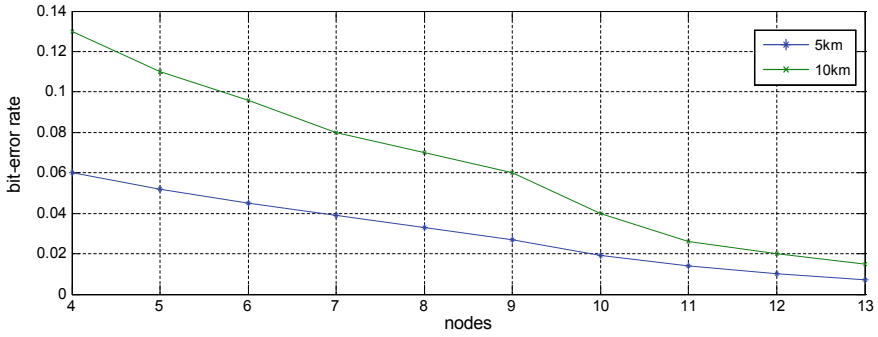


Fig. 6. Bit error rate curve

5 Conclusion

This paper studies the ad hoc network which is applied in underwater acoustic communication, and to obtain each link bandwidth allocation idea of linear programming. According to the transceiver node for increased average SNR, and transceiver node direct communication SNR comparison can be seen from the simulation increase intermediate nodes, improve the signal-to-noise ratio, and reduces the error rate, improve the quality of communication.

Acknowledgment. This work was supported by National Natural Science Foundation of China under Grant 61671144, Shandong Provincial Natural Science Foundation of China under Grant ZR2017BF028, Chizhou University Natural Science Foundation of China under Grant 2016ZRZ006, Anhui Provincial quality engineering project of China under Grant 2014zy077, Chizhou University Teaching and Research Project of China under Grant RE18200030.

References

1. Zhang, B.: Research on acoustic velocity correction algorithm in underwater acoustic positioning. In: Paper Collection of the Ninth Annual China Satellite Navigation Academic Conference—S11 PNT New Concepts, New Methods and New Technologies. China Satellite Navigation System Management Office Academic Exchange Center, vol. 1 (2018)
2. Min, Z., Yan-Bo, W.: Study on the status and prospect of underwater acoustic communication and networking. J. Ocean Technol. **34**(3), 75–78 (2015)
3. Ning, J.: An overview of underwater acoustic communications. Physics **43**(10), 650–657 (2014)
4. Xiaomei, X.: The development and application of underwater acoustic communication and underwater acoustic network. Tech. Acoust. **28**(6), 811–816 (2009)
5. En, C., Fei, Y., Wei, S., Chunxian, G., Wenjun, C., Haixin, S., Xiaoyi, H.: Research progress of underwater acoustic communication technology. J. Xiamen Univ. **50**(2), 271–275 (2011) (Natural Science Edition)
6. Guangpu, Z., Guolong, L., Yan, W., Jin, F.: Integrated system design of distributed underwater navigation, positioning and communication. Acta Armamentarii (12), 1455–1462 (2007)
7. Zeng, W.J., Jiang, X., Li, X.L., et al.: Deconvolution of sparseunderwater acoustic multipath channel with a large time-delay spread. J. Acoust. Soc. Am. **127**(2), 909–919 (2010)
8. Zhong wen, G., Hanjiang, L., et al.: Current progress and research issues in underwater sensor networks. J. Comput. Res. Dev. **47**(3), 37–38 (2010)
9. Xiaoge, L.: Study on the performance of two different modes of underwater acoustic communication network routing protocol. Tech. Autom. Appl. **35**(9), 57–60 (2016)
10. Yunlu, L.: An optimization algorithm of MAC protocol for wireless sensor network. Chin. J. Comput. **35**(3), 529–537 (2012)
11. Angui, W.: Research and improvement of wireless resource allocation algorithm in ad hoc network. Nanjing Univ. Aeronaut. Astronaut. **27**(3), 368–372 (2009)
12. Haiyan, W.: Estimation of SNR in two underwater acoustic channel models. J. Northwest. Polytech. Univ. **27**(3), 368–372 (2009)

Research on the Throughput of Ad Hoc Network with Cyber Encryption

Bo Zhang and Limin Li(✉)

China Academy of Electronics and Information Technology,
Beijing 100041, China
atpcat@163.com

Abstract. Adaptive heuristic for opponent classification (ad hoc) network is characterized by its decentralized structure. All the nodes in ad hoc network are same level and have the ability of message forwarding, which can realize multi-hop communication, with strong expansibility and anti-destructiveness. In avoid of intercepted, deciphered and attacked, cyber security should be emphasized. However, additional bits, which will result in the degeneration of network performance, are needed to be added in media access control (MAC) frame by traditional encryption algorithms. Aiming at this issue, three classical encryption algorithms and their influence on the network throughput is investigated.

Keywords: IEEE 802.11 · Ad hoc · WEP · TKIP · CCMP

1 Introduction

Ad hoc network based on IEEE 802.11 is a special wireless mobile network with the characteristics of decentralized structure, self-organization, self-strengthening, and flexible topology. In ad hoc network, nodes have the ability of message forwarding, which can realize multi-hop communication and increase communication distance dramatically. Multi-hop is the essential difference between ad hoc network and other classical wireless networks with central base station. Communication terminals in ad hoc network can coordinate with each other through hierarchical network protocols, and then realize self-organization communication [1, 2].

Ad hoc network is often deployed to achieve a wide coverage of wireless signal. However, it will result in significant signal leakage in physical space, and then increase the risk of being intercepted, monitored, deciphered, and even attacked in cyber space [3, 4]. Therefore, cyber-security should be emphatically considered to guarantee the availability, confidentiality, integrity and controllability of information when building a wide coverage ad hoc network [5, 6].

The mainstream ad hoc cyber encryption mechanisms require adding redundant information in the MAC frame, which enables the correspondent node obtaining the indication information of encryption and decryption. However, the operation of introducing redundant information into MAC frame will decrease the proportion of

Supported Program: Beijing NOVA Program (No. Z181100006218041).

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 964–973, 2019.
https://doi.org/10.1007/978-981-13-6264-4_114

payload transmission per unit time, further result in the degenerated of network performance. Aiming at this problem, we study the impact of three mainstream encryption algorithms on network throughput in this paper. The throughput of ad hoc is modeled as a two-dimensional Markov process based on Bianchi model [10]. An ad hoc scenario with four nodes is deployed to verify the accuracy of the mathematical model. Finally, the impact of the three encryption algorithms on network throughput is analyzed.

2 The Principle of Cyber Encryption Mechanism

In this paper, wired equivalent privacy (WEP), temporal key integrity protocol (TKIP), the counter and CBC-MAC protocol (CCMP) are considered as the encryption mechanisms for ad hoc network. All the three encryption protocols add redundant information with variant length on the basis of common frames at the MAC layer to implement the function of indicating encryption mechanism and key for correspondent node.

2.1 WEP

In the original version of IEEE 802.11 protocol, the WEP mechanism is defined to enhance the security of transmission. The WEP encryption mechanism is widely used since proposed due to its characteristics of high encrypted data rate and low hardware requirements [7].

The WEP protocol adopts the symmetric encryption algorithm RC4. According to the stream cipher mechanism of the symmetric key, the same plaintext corresponds to the same ciphertext. The initial vector (IV) and integrity check value (ICV) are added to the data frame to prevent unauthorized tampering and bit error caused by data transmission. If WEP encryption is used for communication, the sending node calculates the cyclical redundancy check (CRC) and ICV in the plaintext before sending, and then sends the plaintext and ICV together in ciphertext. The correspondent node first decrypts the ciphertext once receive the message, and then calculates the CRC and ICV of the plaintext. Finally, the CRC and ICV is compared with the original content, and the result is adopted to infer whether the information is tampered during transmission. The MAC layer frame format with the WEP encryption mechanism is shown in Fig. 1(a).

2.2 TKIP

The IEEE 802.11i protocol is released in 2004. It defines the concept of a reliable security network (RSN) [8], and proposes two encryption technologies, which are CCMP and TKIP. As a transition of WEP, TKIP is also based on RC4 symmetric encryption. TKIP protocol enhances the authentication performance and makes several improvements to the WEP defects.

Comparing to WEP, the MAC frame length of TKIP increases by 12 bytes, of which 4 bytes named extended initialization vector (EIV) achieves the function of identifying the encryption mechanism is WEP or TKIP. The other 8 bytes are used to

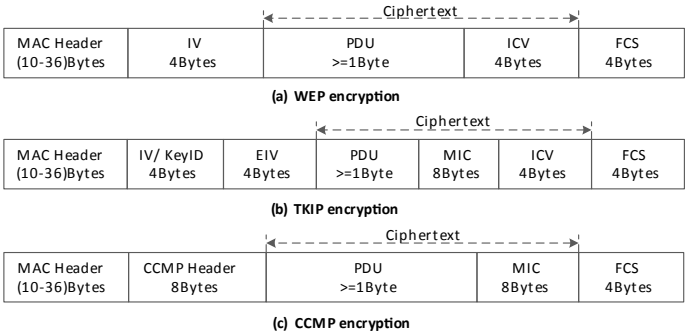


Fig. 1. MAC layer frame format based on three encryption mechanisms

store message integrity code (MIC). If TKIP is adopted as the encryption mechanism for ad hoc network, it calculates the MIC for the plaintext first, and then encrypts the plaintext and the MIC. In short, TKIP improves the security of the MIC according to a high-intensity encryption algorithm. This method reduces the visibility of MAC to the hacker, thereby effectively resisting the fraudulent message. In addition, each MAC frame in TKIP corresponds to a unique TKIP sequence counter (TSC), so that a replay protection mechanism can be established. The MAC layer frame format using the TKIP encryption mechanism is shown in Fig. 1(b).

2.3 CCMP

CCMP is the mandatory encryption method specified in the IEEE 802.11i protocol, and it adopts advanced encryption standard (AES), which can resist replay analysis attack and guarantee the integrity of information.

CCMP added 8-byte frame header and 8-byte MIC to MAC frame. According to the general block encryption authentication mode, CCMP needs to specify two parameters for the length domain and the authentication domain, and to balance the security and computational overhead. The length domain is two 8-bit code blocks, and the authentication domain is eight 8-bit code blocks. In addition, CCMP requires each session to adopt a new temporary password, with different serial Numbers for each message, which can resist replay attacks and password analysis attacks. The CCMP password protocol takes AES as the core, and the encryption and authentication functions are respectively implemented by the combination of the counter mode (CTR) and CBC-MAC mode [9]. The MAC layer frame format with CCMP encryption mechanism is shown in Fig. 1(c).

3 Mathematical Model of Ad hoc Network Throughput Performance

Based on the Bianchi model [10], this section establishes a throughput performance model based on cyber-encrypted ad hoc communication network. The Bianchi model needs to meet the prerequisites: all terminals work in saturated state, that is, there is always at least one transfer packet at any time; the wireless channel is in the ideal state, which means the transmission error caused by the channel noise will not occurred, but the packet collision is not excluded; there is no hidden terminal in the network, so that all communication terminals can listen to the transmission state of the channel.

Given that there are n terminals in the ad hoc scenario, the collision probability of each terminal sending packet is p, and the probability of collision is independent of each other. For each terminal, the random process $b(t)$ represents the value of the back-off counter at the time of t, and the unit of t is the back-off slot. Assuming the maximum length of contention window and the minimum length of contention window are CW_{max} and CW_{min}, respectively. Therefore, we can obtain $CW_{min} = W$ and $CW_{max} = 2^m W$, whereas m represents back-off order. Obviously, the back-off order at time t is a stochastic process. On this basis, define $\{s(t), b(t)\}$ as a two-dimensional Markov process. For simplicity, we use $Pr\{i_1, k_1 | i_0, k_0\}$ to represent $Pr\{s(t+1) = i_1, b(t+1) = k_1 | s(t) = i_0, b(t) = k_0\}$, which is the probability of Markov state transition in the rest of the paper. The state transition process is shown in Fig. 2.

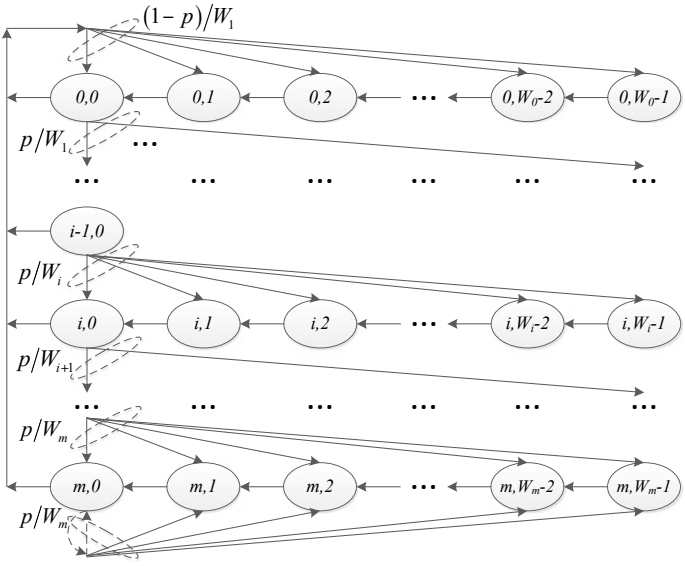


Fig. 2. Two-d Markov model for IEEE 802.11 DCF protocol

In the Markov model shown in Fig. 2, non-null one-step transition probabilities are shown below:

A. During idle timeslot, the back-off counter is decremented.

$$\Pr\{i,k|i,k+1\}=1 \quad k\in(0,W_i-2) \quad i\in(0,m) \tag{1}$$

B. Following a successful packet transmission, a new packet transmitting starts with back-off stage 0, and thus the back-off is reset in the range (0,W_0-1) uniformly.

$$\Pr\{0,k|i,0\}=(1-p)/W_0 \quad k\in(0,W_0-1) \quad i\in(0,m) \tag{2}$$

C. After an collision occurs at back-off stage $(i-1)$, the back-off stage increases, and the new back-off value is reset in the range (0,W_i) uniformly.

$$\Pr\{i,k|i-1,0\}=p/W_i \quad k\in(0,W_i-1) \quad i\in(1,m) \tag{3}$$

D. Once an collision occurs at back-off stage m, it will not increase in the following transmissions.

$$\Pr\{m,k|m,0\}=p/W_m \quad k\in(0,W_m-1) \tag{4}$$

Considering the Markov chain approaching to the stable state, and let $b_{i,k}$ be the stationary distribution of the chain, which satisfies

$$b_{i,k}=\lim_{t\to\infty}\Pr\{s(t)=i,b(t)=k\},k\in[0,W_i-1],i\in[0,m] \tag{5}$$

For $k\in(1,W_i-1)$, we can derive

$$b_{i,k}=\begin{cases}\left(\frac{W_i-k}{W_i}\right)(1-p)\sum_{j=0}^{m}b_{j,0}, & i=0\\ \left(\frac{W_i-k}{W_i}\right)pb_{i-1,0}, & 0<i<m\\ \left(\frac{W_i-k}{W_i}\right)p\left(b_{m-1,0}+b_{m,0}\right), & i=m\end{cases} \tag{6}$$

Define τ as the transmission probability of any terminal at any time slot. Considering that the trigger condition of the terminal sending packet is that its back-off counter reaches zero, then we can obtain the following expression referring to [10]

$$\begin{cases} p = 1 - (1-\tau)^{n-1} \\ \tau = \frac{2(1-2p)}{(1-2p)(W+1)+pW(1-(2p)^m)} \end{cases} \tag{7}$$

Let P_{tr} be the probability that there is at least one terminal to send a packet at a particular slot. The probability P_s represents that a transmission is successful. It means that exactly one station transmission occurring on the channel at the time slot. We can obtain the relationship as below

$$\begin{cases} P_{tr} = 1 - (1-\tau)^n \\ P_s = \frac{n\tau(1-\tau)^{n-1}}{P_{tr}} = \frac{n\tau(1-\tau)^{n-1}}{1-(1-\tau)^n} \end{cases} \tag{8}$$

Supposing S is the overall network throughput, let E[l] be the average length of the packet, and then S can be expressed as follows:

$$\begin{cases} S = P_s P_{tr} \mathrm{E}[l] \cdot (\mathrm{E}[\sigma_{slot}] + \mathrm{E}[T_s] + \mathrm{E}[T_c])^{-1} \\ \mathrm{E}[\sigma_{slot}] = (1 - P_{tr})\sigma_{slot} \\ \mathrm{E}[T_s] = P_{tr} P_s T_s \\ \mathrm{E}[T_c] = P_{tr}(1 - P_s) T_c \end{cases} \tag{9}$$

In Eq. (10), σ_{Slot} represents back-off interval, T_s represents the channel duration of single successful packet transmission, T_c represents the channel duration of single collide packet transmission. The distance between two terminals could be very long in a large scene cyber-encrypted ad hoc network, thus RTS/CTS should be adopted in MAC protocol in avoid of "hidden terminal" problem. T_s and T_c can be spread as follow:

$$\begin{cases} T_s = RTS + 3SIFS + DIFS + CTS + ACK + H + E[P] + 4\delta \\ T_c = RTS + DIFS + \delta \end{cases}, \tag{10}$$

where ACK indicates the acknowledgement frame, H represents PHY and MAC layer overhead, which includes indication overhead for communication and encryption overhead, and each parameter unit needs to be converted to length of time.

Supposing n terminals exist in the cyber-encrypted ad hoc network, in which the number of single-hop link and the number of two-hop link are Sh_1 and Sh_2, respectively. Assuming that multi-hop excess two does not exist, therefore E[T_s] can be generalized as

$$\mathrm{E}[T_s] = P_{tr} P_s \left(\frac{Sh_1 T_s^{(1)} + Sh_2 T_s^{(2)}}{Sh_1 + Sh_2} \right), \tag{11}$$

where $T_s^{(1)}$ and $T_s^{(2)}$ represent the packet transmission time of cyber-encrypted single-hop and two-hop link.

4 Simulation and Performance Evaluation

In this section, we use Monte Carlo simulation method to validate the precision of the mathematical model derived in this paper. The simulation scenario aims at a cyber-encrypted ad hoc network including four terminals. On this basis, the influence of three cyber encryption algorithms on network performance is investigated. The ad hoc network model is shown in Fig. 3, including eight single-hop links, four two-hop links, and dotted lines indicating the range of signal radiation. The related simulation parameters are shown in Table 1.

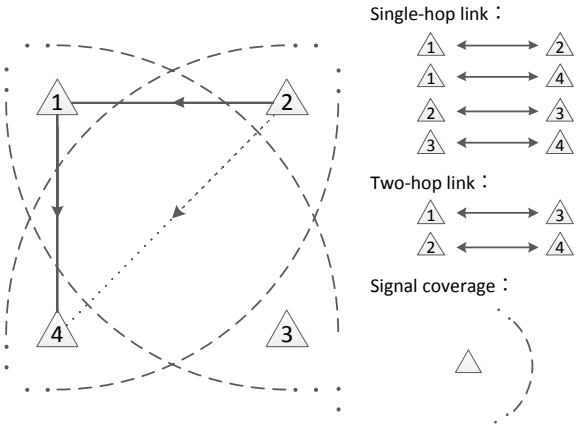


Fig. 3. Cyber-encrypted ad hoc network model

Table 1. Simulation parameters

Parameter	Value
Encryption methods	WEP/TKIP/CCMP
MAC protocol	IEEE 802.11a DCF RTS/CTS
MCS scheme	BPSK + 1/2 coding rate
Channel spacing	5 MHz
Minimum/Maximum back-off window size $(\boldsymbol{W}/\boldsymbol{2^m W})$	15/1023
RTS/CTS	160/112 bits
SIFS/DIFS	32/58 μs
MAC header	224 bits
Packet payload	256/512/1024 bits
Slot time	13 μs
ACK frame length	112 bits
PLCP header	16 μs
PLCP preamble	64 μs

Simulation results are shown in Figs. 4, 5, 6 and 7. Base on the encryption methods of WEP, TKIP, CCMP, and none-encryption, Figs. 4, 5 and 6 demonstrates the

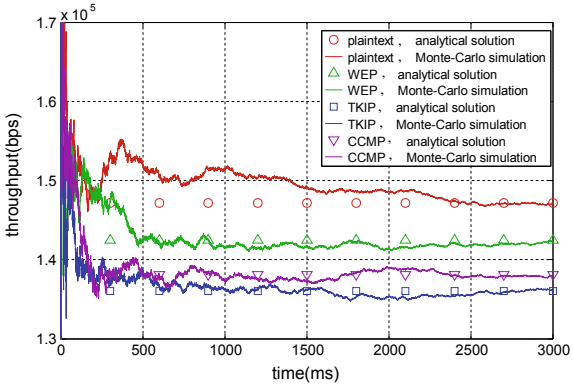


Fig. 4. Network throughput base on cyber encryption ($E[l]$ = 256 bit)

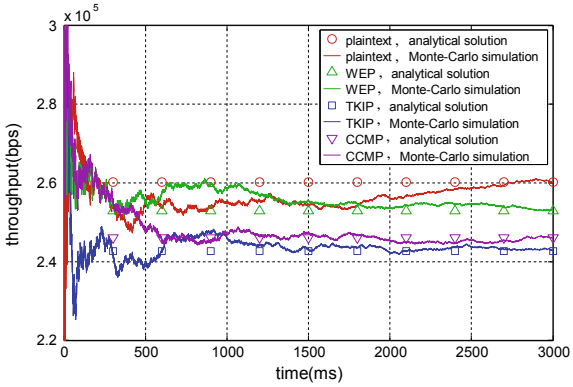


Fig. 5. Network throughput base on cyber encryption ($E[l]$ = 512 bit)

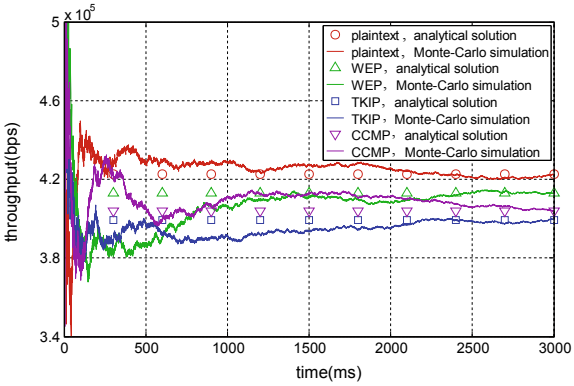


Fig. 6. Network throughput base on cyber encryption ($E[l]$ = 1024 bit)

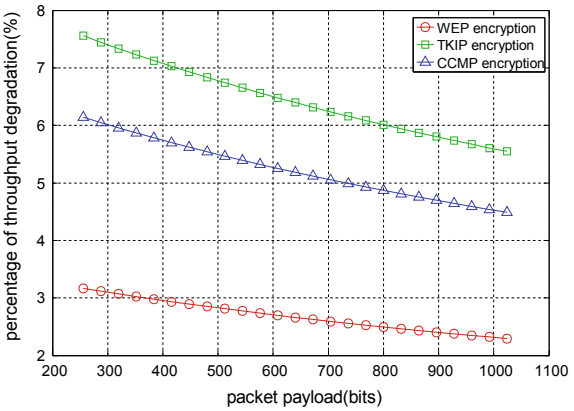


Fig. 7. Throughput influence caused by cyber encryption

network throughput transmitting with packet payload 256 bits, 512 bits, and 1024 bits, respectively. The theoretical results and Monte Carlo simulation results are presented correspondingly in the three group simulations. It can be seen that, along with the transmitting, the network throughput based on Monte Carlo simulation gradually converges to the theoretical results presented in the previous section. Among the three encryption methods, best network performance can be obtained with WEP, and the network performance with TKIP is the worst, while network performance with CCMP is situated between them. The result is due to the fact that the three encryption algorithms needs different lengths of redundant information to be added in the MAC layer, therefore the network throughput is decreased correspondingly compared to plaintext transmission. Figure 7 indicates the throughput degradation percentage of the three encryption methods along with various packet payloads comparing to the plaintext transmission. Simulation results indicate that, with the packet payload length $E[l] = 512$ bit, the network throughput adopting WEP, TKIP and CCMP encryption methods decreased by 2.8%, 6.7% and 5.5%, respectively.

5 Conclusions

Aiming at ad hoc communication network technology, this paper investigates the influence of three major cyber encryption methods on network performance. The result shows that network throughput decreases linear proportionally to the length of redundant information added in MAC layer approximately. According to mathematical modeling, the analytical solutions of the influence on network throughput with the three cyber encryption methods are presented, and the accuracy is verified by Monte Carlo simulation.

References

1. Zhang, Y., Song, L., Jiang, C., Tran, N.H., Dawy, Z., Han, Z.: A social-aware framework for efficient information dissemination in wireless ad hoc networks. IEEE Commun. Mag. **55**(1), 174–179 (2017)
2. Seliem, H., Ahmed, M.H., Shahidi, R., Shehata, M.S.: Delay analysis for drone-based vehicular ad-hoc networks. In: 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), pp. 1–7 (2017)
3. Ghica, O.C., Nita-Rotaru, C., Trajcevski, G., Scheuermann, P.: Security of electrostatic field persistent routing: attacks and defense mechanisms. Ad Hoc Netw. **36**, 270–295 (2016)
4. Granjal, J., Monteiro, E., Sa Silva, J.: Security in the integration of low-power wireless sensor networks with the internet: a survey. Ad Hoc Netw. **24**, 264–287 (2015)
5. Korba, A.A., Nafaa, M., Ghanemi, S.: An efficient intrusion detection and prevention framework for ad hoc networks. Inf. Comput. Secur. **24**(4), 298–325 (2016)
6. Wu, W.: The research of cyberspace and communication network security problems. J. China Acad. Electron. Inf. Technol. **6**(5), 473–476 (2011)
7. Wang, L., Wang, L., Zhang, Y.: Analysis of WEP and WEP2 algorithm security in wireless LAN. Comput. Eng. **30**(12), 111–113 (2014)
8. IEEE Std. 802.11i Part11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Amendment 6: Medium Access Control (MAC) Security Enhancements. IEEE Computer Society, New York, USA (2004)
9. Wu, Y.: Application and Analysis in IEEE 802.11i based Security of Wireless Network. Shanghai Jiao Tong University, pp. 18–21 (2008)
10. Bianchi, G.: Performance analysis of the IEEE 802.11 distributed coordination function. IEEE J. Sel. Areas Commun. **18**(3), 535–547 (2000)

An Efficient Algorithm for Invulnerability Enhancement in Cognitive Ad Hoc Networks

Xiao Zhang(✉), Yukun Wang, Yueliang Liu, and Wenbin Guo

School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing 100876, China
pao-pao@bupt.edu.cn

Abstract. The communication network often loses connections due to the energy consumption of the equipments, the failures of hardware, and malignant attacks. Therefore, the research of the invulnerable network has important theoretical value and practical significance. On account of the multiple-tree infrastructure of current communication networks, a graph-based model is established for constructing invulnerable network. Specifically, network structure entropy is put forward as the global evaluation metric, and an optimization algorithm satisfying the minimum overhead of the invulnerable constraint is proposed. Simulation results show that the proposed algorithm significantly reduces network overhead while considering network invulnerability compared with ILP linear programming algorithm.

Keywords: Network invulnerability · Connectivity · Network structure entropy · Networking cost

1 Introduction

Network invulnerability refers to the ability of network to maintain normal function under the circumstances of external attack or internal damage, which represents the insensitivity of network function to internal or external interference [1]. For most networks, such as power systems, logistics systems, and traffic systems, a few key links or nodes are destroyed, the entire network operation will be affected, even paralyzed, leading to huge losses. According to the researches of invulnerability networking, studies in [2, 3] propose the evaluation metrics that is mainly considered from connectivity, the maximum connected branch ratio, the shortest path, and the node removal ratio. From the perspective of network maintenance capabilities, the authors in [4] define the efficiency of network communication, and combined with connectivity to evaluate the invulnerability of network, which is more comprehensive. While the network invulnerability increases the network's ability to maintain normal function, it demands extra network overhead. Furthermore, the rapid expansion of the demand for spectrum resources leads to the desire for low cost, extensible, and invulnerable networks [5].

Network topology ensures the reliability of the network, depending on the nodes being connected. The topology of 2-degree connectivity networks without parallel edges is proved to be a NP-hard problem [6]. The ILP linear programming model

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 974–982, 2019.
https://doi.org/10.1007/978-981-13-6264-4_115

proposes a reasonable system overhead scheme for the 2-degree connectivity for small-scale networks. In a relatively short time, the overhead can be kept in a certain range, but with the expansion of the network scale, the cost is still large [7, 8]. The heuristic topology design and the reverse decomposition method are used to analyze the total number of links to minimize the network, but it is not explicitly proposed for how to implement the least link resistant network [9].

In this paper, an invulnerable network algorithm is proposed based on the tree structure, which can ensure the connectivity of the network when any node disappears due to its own damage or external causes. First, the paper establishes a graph-based model according to real ad hoc networks. Under the connectivity constraint, we proposed an optimization formulation to satisfy the minimum overhead. Then an optimization networking algorithm is presented, and the theoretical analysis is given. In addition, network structure entropy is put forward as a global metric to evaluate the network invulnerability. Simulation results show that compared with the algorithm based on ILP model for low-level nodes to connect to advanced nodes, there is a significant improvement in the network performance, such as network overhead and networking success rate.

2 Problem Description

2.1 Graph-Based Model

Consider a network $G(V, E, \omega)$ with n nodes to be connected. The node set is expressed as $V = \{v_1, v_2, \ldots, v_n\}$, and the edge set of the link between nodes is $E = \{e_{ij} | e_{ij} = (v_i, v_j), v_i, v_j \in V\}$. The overhead brought by the link establishment is set as the weight set $\omega = \{\omega_{ij} | \omega_{ij} = \omega(v_i, v_j), v_i, v_j \in V\}$.

Network overhead refers to resources to be used in the process of network establishment. In this paper, we mainly consider energy consumption. For communication systems, the energy consumed by establishing links between nodes is defined as [10]

$$\begin{aligned} E_{Tx}(k,d) &= E_{Tx-elec}(k) + E_{Tx-amp}(k,d) \\ &= \begin{cases} kE_{elec} + k\varepsilon_{fs}d^2, & d < d_0 \\ kE_{elec} + k\varepsilon_{mp}d^4, & d \geq d_0 \end{cases} \end{aligned} \tag{1}$$

where d represents the distance between two nodes, transmits information of k bits. ε and d_0 denotes the amplification factor and the threshold of the distance, respectively. Based on the network model mentioned above, we define

$$a_{ij} = \begin{cases} 0, & e_{ij} = \phi \\ 1, & e_{ij} \neq \phi \end{cases}, (i,j) \in V \tag{2}$$

$$\omega_{ij} = \begin{cases} E_0 + \varepsilon_{fs} \times d_{ij}^2, d_{ij} < d_0 \\ E_0 + \varepsilon_{mp} \times d_{ij}^4, d_{ij} > d_0 \end{cases}, (i,j) \in V \tag{3}$$

where $A(G) = \{a_{ij}\}$ represents the corresponding adjacency matrix of the system, and $\omega(G) = \{\omega_{ij}\}$ is the cost matrix of the system.

2.2 Problem Formulation

The connectivity of a network refers to whether there are any paths between any two nodes in the network. A network's connectivity can be defined by $K(G)$, which means the minimal number of $K(G)$ nodes that can make the whole network not connected. A graph with connectivity of no less than K is called a K-connected graph. The purpose of this paper is to design a network structure with minimal overhead under the constraint of network connectivity. In view of the constraints on the network in the ILP model, the connectivity of the same constraint network is at least 2.

The system overhead can be expressed as the sum of the node point multiplication between the cost matrix and the adjacency matrix [11]. Based on the above analysis, the invulnerability network model (HILP) built in this paper is designed as follows:

$$\begin{aligned} \min Z &= \sum_{(i,j)\in V} a_{ij}\omega_{ij} \\ s.t. \quad & K \geq 2 \end{aligned} \tag{4}$$

where K denotes the connectivity of the network, in other words, it is a K-connected graph.

2.3 Metric of Network Invulnerability

The connectivity of the nodes reflects the importance of nodes in the network, such as the few "core nodes" with a large number of connections in complex networks and a large number of "terminal nodes" with very few connections, which also known as the non-homogeneity nodes. Using λ_i to represent the degree of node v_i, then the definition of the importance of the node is given as $I_i = \lambda_i / \sum \lambda_i$. The network structure entropy is the measure of the system homogeneity [12]. If the network is connected randomly and the node is very important, then the network has a strong homogeneity, otherwise, the network is not homogeneous. The definition of network structure entropy H is

$$H = -\sum_{i=1}^{n} I_i \ln I_i \tag{5}$$

When the nodes in the network have the same importance, the network has the strong homogeneity, and the invulnerability is the strongest. In this case, $I = 1/n$. The node degree is the same, and the network structure entropy takes the maximum value $\ln n$.

3 Invulnerability Networking Algorithm

We construct an invulnerable network satisfying 2-degree connectivity based on balanced multi-tree edge adding. It is known from graph theory that the connection degree of a node can be improved effectively by adding edge between nodes. But at the same time, the system energy consumption will increase. Therefore, the objective of the optimization is to add fewer edges to increase the network invulnerability as much as possible, while satisfying the connectivity constraint.

According to the graph theory, for all the tree structures, except the leaf nodes, the rest are all cut points (hereinafter referred to as the superior nodes). If the tree structure is to be transformed into a 2-degree connected network, all the cutting points should be eliminated first. Eventually, any node in the network is in a circle. This way is called the homogenization.

Based on the optimal design of the original system structure, as shown in Fig. 1 (left), it is assumed that the initial network structure is a single symmetric balance tree, from top to bottom are, in order, the root node, the intermediate nodes, and the leaf nodes. The leaf node selects the leaves with the least distance within the scope of its communication. If there is still a leaf node with degree of 1, then the node is connected to the leaf node with the smallest communication distance and the degree of 2. The network structure after the edge adding algorithm is shown in Fig. 1 (right).

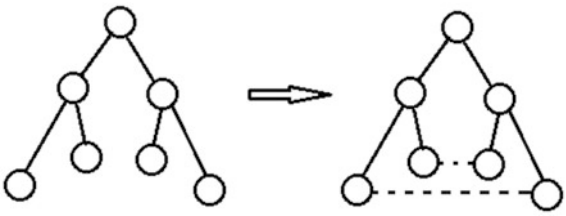


Fig. 1. Homogeneity process of single root tree

- *Property 1 (Minimum edges):* The number of adding edges is least by the way of homogeneity adding.

Proof: According to graph theory, for any network structure, if there is a leaf node whose degree is 1, then the upper node of the leaf node must be cut point. Figure 1. (right) is a 2-degree connected graph. As the added edges are all connected to the leaf node, if any edge is removed, at least one degree of the leaf node is 1. Then there is a cut point in the graph. So the number of edges added to the 2-degree connected graph based on tree optimization algorithm is the least.

- *Property 2 (2-connectivity guarantee):* By using the homogenization method, there is no cut point in the connected graph, that is, the connectivity of the topology is at least 2.

Proof: For tree structure, all leaf nodes are not cut points, and the optimized leaf nodes are still not cut points. It is assumed that there is a cut point in the network with a higher node v, that is, after removing the node v and its connected side, there are two separated network structures, G_1 and G_2, and there are two nodes u and w in G_1 and G_2, respectively. When v is an intermediate node, the two nodes, u and w, are their own leaf nodes. As any leaf node is connected to the leaf nodes under other intermediate nodes, the other intermediate nodes are connected by root nodes and are in one network, which is contradictory to the hypothesis. When v is the root node, the two nodes u and w are both intermediate nodes. If u and w are in the two branches, the corresponding leaf nodes are also in the two branches, apparently not.

For a multiple roots symmetric balance tree, as shown in Fig. 2 (left), which is connected by multiple root nodes in the network, first connects the leaves of each root to a homogeneity, and then regards the intermediate node as a leaf node and still is connected in a homogeneity way. As shown in Fig. 2 (right), the network is still a 2-degree connected graph with the least adding edges.

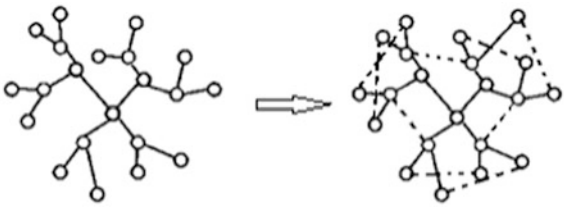


Fig. 2. Homogeneity process of multiple roots tree

Based on the above analysis, assuming the number of leaf nodes is m, and the number of intermediate nodes is w. First, the leaf node selects the leaves with the least distance within the scope of its communication, the two nodes that will be connected should not belong to the same intermediate node. Leaf nodes run the procedure one by one until there is no leaf node in the network. Then the intermediate node selects peer to peer as the leaf node do until it comes to the root. The detailed process of the algorithm is as follows:

Algorithm **HILP**

1. **Input**: Adjacency matrix $A(G)$, node position set P of symmetric balanced tree. Leaf nodes $ID_{leaf} = i,\ i \in [1, m]$, intermediate nodes $ID_{int} = p,\ p \in [1, w]$.
2. **Initialization**: $i = 1, p = 1$.
3. **While** $i \le m$ do Give neighbors $ID_i = j$ by distance, initialize $j = i + 1$
   While $j \le m$ do
   If $K_i < 2, K_j < 2$, and v_i, v_j do not belong to the same intermediate node
   Then add edge $e_{ij} = (v_i, v_j)$, update $A(G)$, break out of the current loop
   Else $j = j + 1$
   End while, If $K_i < 2$ Then $j = i + 1$
   While $j < m$ and v_i, v_j belong to the same intermediate node
   do $j = j + 1$ End while , Add edge $e_{ij} = (v_i, v_j)$,update $A(G)$
   $i = i + 1$
   End
4. Similar operations to intermediate nodes to step 3.
5. **Return** the current adjacency matrix $A(G)$ and the target value Z.

It is unavoidable in cognitive ad hoc networks that there is one leaf node which cannot find any unconnected node within the communication scope. In this case, we can improve the invulnerability of the link by improving the quality of the unique connection and increasing the number of available channels.

4 Simulation Results

4.1 Comparison and Analysis of Algorithms

A symmetric balance tree structure is established for the different number of nodes in the network. The ILP model algorithm and HILP model algorithm are used to obtain the 2-degree connectivity invulnerable network in the following simulation. The comparison of the performance metrics including the number of required additional edges, the changes of the network structure entropy and the network overhead that obtained by the two algorithms is as follows.

It is known from Fig. 3 (left) that when the number of nodes is simultaneous, if the network satisfies 2-degree connectivity, the number of edges added to the ILP algorithm is much larger than that of the HILP algorithm. When the number of network nodes is the same, if the network satisfies the 2-degree connectivity, the number of edges added by the ILP adding algorithm is obviously larger than that of the HILP adding algorithm. With the increase of the number of the nodes in the network, the number of edges added by ILP is increased by two times that of HILP. With the increase of the number of edges added, the network overhead is increasing. As is shown in Fig. 4 (right), the network overhead of HILP is obviously less than that of ILP.

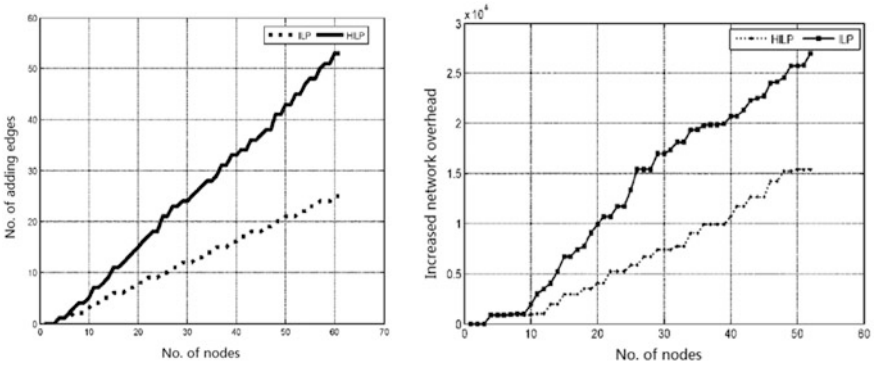


Fig. 3. Network overhead caused by invulnerability networking

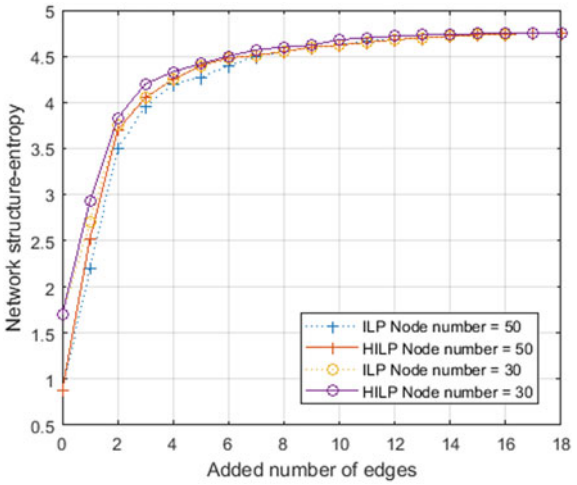


Fig. 4. The influence of the added number of edges on the network structure entropy

Network structure entropy increases with the decrease of non-homogeneity among nodes, and the network invulnerability is better. When the number of edges is increasing, the importance of nodes is becoming similar, and the structure entropy increases slowly. In small-scale networks, the initial network structure entropy is relatively large because of the little difference in node degree. From Fig. 4, we can find that the network structure entropy in the HILP algorithm, which adds edges between peers, increases faster than that of ILP algorithm, which adding edge to the superior nodes.

4.2 Algorithm Feasibility Analysis

Communication devices in communication networks are affected by the transmission power and cannot communicate with devices outside the communication range. In order to verify the feasibility of the algorithm in the communication network, the simulation results show that when the number of nodes is different, the success rate of invulnerability networking algorithm change with the increase of communication radius. The distribution range of nodes is 100 m * 100 m. As shown in Fig. 5, the less the number of nodes in the network, the greater the communication radius of the network success rate up to 100%, that is, the smaller the constraints of the communication range to the success rate of the network. When the nodes in the network are sparse, and the communication radius R is about 30 m, the success rate of networking is nearly 100%. This shows that the algorithm has high feasibility.

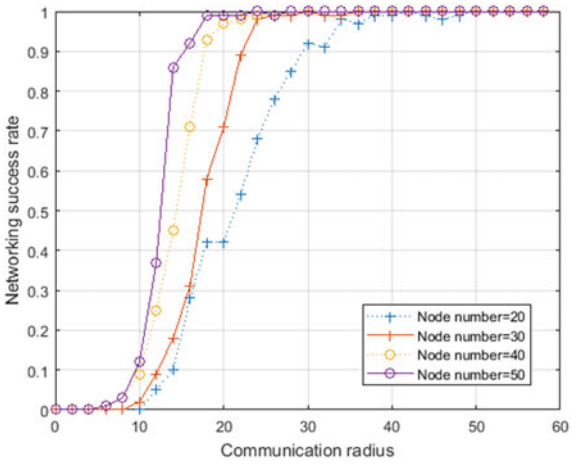


Fig. 5. The influence of communication radius on the networking success rate

5 Conclusion

In this paper, a graph-based model is established according to the real ad hoc networks. Under the connectivity constraint, we proposed an optimization formulation to satisfy the minimum overhead. Then an invulnerability networking algorithm with less network overhead to satisfy the 2-degree connectivity is proposed based on the tree structure. The impact of network structure entropy on the homogeneity of nodes is analyzed. Compared with the previous ILP model algorithm, there is a significant improvement in the overhead of networking. Simulation analysis shows that the proposed invulnerability networking algorithm is effectiveness. In future study, we will analyze the invulnerability of network connectivity of more complex networks.

References

1. Wu, R.Z., Zou, Y.J., Fan, B.: Research on network invulnerability based on DWS model considering cascading failure in power data network. In: International Conference on Communication Technology (ICCT), pp. 1233–1237 (2017)
2. Frank, H., Frisch, I.: Analysis and design of survivable networks. IEEE Trans. Commun. Technol. **18**(5), 501–519 (1970)
3. Crucitti, P., Latora, V., Marchiori, M., et al.: Error and attack tolerance of complex networks. Nature **340**(1–3), 388–394 (2004)
4. Crucitti, P., Latora, V., Rapisarda, A., et al.: Efficiency of scale-free networks. Phys. A Stat. Mech. Appl. **320** (2002)
5. Jothi, R., Raghavachari, B.: Survivable network design: the capacitated minimum spanning network problem. Inf. Process. Lett. **91**(4), 183–190 (2004)
6. Kerivin, H., Mahjoub, A.R.: Design of survivable networks: a survey **46**(1), 1–21 (2005)
7. Soni, S., Pirkul, H.: Design of survivable networks with connectivity requirements. Telecommun. Syst. **20**(1), 133–149 (2002)
8. Balakrishnan, A., Mirchandani, P., Natarajan, H.P.: Connectivity upgrade models for survivable network design. Oper. Res. **57**(1), 170–186 (2009)
9. Chaves, D.A., Bastosfilho, C.J., Martinsfilho, J.F.: Multiobjective physical topology design of all-optical networks considering QoS and Capex. Opt. Fiber Commun. 1–3 (2010)
10. Heinzelman, W.R., et al.: Energy-efficient communication protocol for wireless microsensor networks. In: Adhoc & Sensor Wireless Networks, vol. 2, p. 10 (2000)
11. He, J.H., Fan, Y.Z., Li, S.D.: The spectrum of a graph obtained from copies of a generalized bethe tree and a complete graph. Chin. J. Eng. Math. **29**(5) (2012)
12. Tan, Y.J., Wu, J.: Network structure entropy and its application in scale-free networks. Theory Pract. Syst. Eng. **24**(6), 1–3 (2004)

Multiuser MIMO Relay System Model and Performance Evaluation Under Hardware Impairment

Ruiqing Zhang, Hui Li(✉), Ming Ye, and Yafei Yang

College of Information Science and Technology, Hainan University, Haikou, China
{iezhangruiqing, hitlihuillll2}@163.com, 2248817407@qq.com, yafei94@yeah.net

Abstract. The multiuser MIMO system uses multiple transceiver antennas, which can fully develop space resources. Without increasing the spectrum resources and transmitting power of the antenna, the capacity of the system can be increased. MIMO technology can not only improve the capacity of the system, but also improve the reliability of wireless transmission and reduce the bit error rate. In this paper, we study the influence of MIMO relay system with antenna hardware damage on the information transmission between multiple users. And analyze the capacity and bit error rate of the system. Second, two commonly used signal detection methods are adopted in uplinks and downlinks. They are zero forcing algorithm and minimum mean square error algorithm respectively. Compared with ideal antenna MIMO relay system, the influence of antenna damage on MIMO relay station cannot be ignored. In the theoretical analysis, hardware damage model is more conducive to the reference and application in practical engineering.

Keywords: MIMO · Hardware impairment · Signal detection · Bit error rate

1 Introduction

Since the advent of wireless communications, data traffic in wireless communications has almost doubled every two and a half years. Especially in recent years, data traffic is increasing exponentially with the rising of mobile Internet. MIMO (Multiple input and multiple output) technology is one of the key technologies of the fifth generation mobile communication system. It can obviously improve the performance of wireless communication system in terms of system capacity, spectrum efficiency, and strong anti-interference capability. According to the principle of conservation of energy, the more energy the fraction is, the smaller the energy of each part is. For the transmitting system, the transmission power of each antenna decreases with the increase of their number when the total transmit power is constant. Therefore, increasing the number of antennas will become the trend and direction of MIMO in the future. If the number of antennas is large enough, its hardware performance will directly affect the performance of MIMO system. Therefore, when analyzing the multiuser MIMO relay system model, we need to take into account the hardware impairment of the antennas.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 983–991, 2019.
https://doi.org/10.1007/978-981-13-6264-4_116

Gustavsson et al. considered the impact of hardware damage on the multiuser MIMO base station system under Rayleigh channel [1]. In [2], Sun et al. considered the performance of multi relay MIMO systems with the hardware damage under Rician channel. According to the analysis of the existing literature, there is no comparison between the research and analysis of the performance of the hardware damage to the multiuser MIMO relay system under the Rayleigh channel and the Rician channel. In this paper, the hardware damage of MIMO relay antenna is considered. And its modeling is analyzed under different channels. In the signal detection of uplink and downlink, common zero forcing detection algorithm and minimum mean square error algorithm are adopted. The system error rate is analyzed with the increase of antenna number or the increase of signal-to-noise ratio. This paper first introduces the multiuser MIMO relay system model, and then introduces the signal detection methods used in the up and down links. Finally, the results of the simulation and the related conclusions are given.

The symbols of this article are as follows. $(\cdot)^{\mathrm{T}}$, $(\cdot)^{*}$, $(\cdot)^{\mathrm{H}}$, $(\cdot)^{-1}$, tr$(\cdot)$ represent the transpose, conjugate, conjugate transposition, inverse and trace of the matrix respectively. $[\cdot]_{ij}$ is the i row and j column element of the matrix. $E[\cdot]$ represents the expected operation.

2 System Model

We established a system with $2L$ users. The information transmission process between users is carried on by a relay station. The number of antennas at the relay station is M, and the number of antennas per user is N. Information transmission includes the uplink transmission and the downlink transmission respectively [3]. In the process of uplink transmission, $2L$ users send signals to relay at the same time and relay receive signals. During downlink transmission, the relay amplifies the received signal and sends the amplified signal to all users. $\boldsymbol{H}_i(i = 1, 2, 3 \cdots 2L)$ represents the uplink matrix of the user to the relay. Here, we set the uplink and the downlink to use the same channel. The downlink matrix from relay to user can be expressed as $\boldsymbol{H}_i^{\mathrm{H}}(i = 1, 2, 3 \cdots 2L)$, which is conjugate transpose matrix. The system model is shown in Fig. 1.

In the uplink model [4], $\boldsymbol{u}_k$ is the signal sent by the user k, and the $\boldsymbol{v}$ is the signal received by the relay. $\boldsymbol{n}_R$ is the noise vector at the relay, which satisfies the cyclic symmetric complex Gaussian distribution and has the unit covariance matrix. That is $\boldsymbol{n}_R \sim CN(0, \boldsymbol{I})$. The hardware damage model of the receiver at the relay can be built to add an independent zero mean Gaussian noise on each antenna, and the variance is proportional to the power of the receiving signal. In this way, the distortion noise at the receiving end can be obtained as $\boldsymbol{r}_R$, satisfying the relation of $\boldsymbol{r}_R \sim CN(0, \xi_R \mathrm{diag}(E[\boldsymbol{v}_R \boldsymbol{v}_R^{\mathrm{H}}]))$ [5].

The relay receives the signal

$$\tilde{\boldsymbol{v}}_R = \boldsymbol{H}_i \boldsymbol{U}_k + \boldsymbol{n}_R + \boldsymbol{r}_R, \tag{1}$$

in which, $\boldsymbol{H} = [\boldsymbol{H}_1, \cdots, \boldsymbol{H}_{2L}]$, $\boldsymbol{U}_k = [\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_{2k-1}, \boldsymbol{u}_{2k}]^{\mathrm{T}}$.

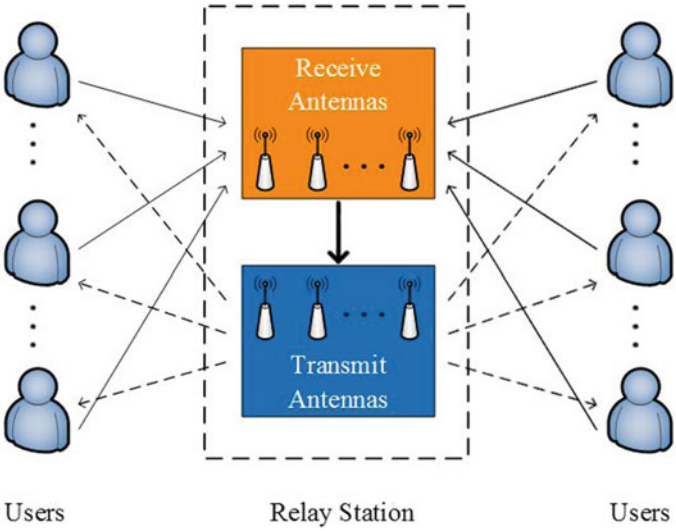


Fig. 1. System model

The limited power of each user is P_k in the uplink. If the user k transmit covariance matrix is $\boldsymbol{R}_{kk} = E[\boldsymbol{u}_k \boldsymbol{u}_k^{\mathrm{H}}]$, the power constrained condition can be expressed as $\mathrm{tr}(R_{kk}) \leq P_k \, k = 1, \cdots, k$.

In the downlink model [4], the signal sent from the relay station is $\boldsymbol{X}$, and the signal received by the mobile user k is $\boldsymbol{y}_k$. $\boldsymbol{n}_k$ is the noise vector at the user k, which satisfies the cyclic symmetric complex Gaussian distribution and has a unit covariance matrix, which is $\boldsymbol{n}_k \sim CN(0, \boldsymbol{I})$. The hardware damage of the receiver, an independent zero mean Gauss noise is added to the antenna on the receiver of the user k, then the noise $\boldsymbol{r}_k$ of the user k receiver satisfies the relationship of $\boldsymbol{r}_k \sim CN(0, \xi_k E[\boldsymbol{y}_k \boldsymbol{y}_k^{\mathrm{H}}])$ [5].

The signal received by user k is

$$\tilde{\boldsymbol{y}}_k = \boldsymbol{H}_i^{\mathrm{H}} \boldsymbol{X} + \boldsymbol{n}_k + \boldsymbol{r}_k. \tag{2}$$

In (2), the covariance matrix of the transmitted signal is $\boldsymbol{R}_{xx} = E[\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}]$. If the average power of the relay is P, the power constrained condition can be expressed as $\mathrm{tr}(\boldsymbol{R}_{xx}) \leq P$.

3 Signal Detection Method

In essence, the fundamental purpose of signal detection is to restore the original sending signal as much as possible through the received signal and a known channel information. In the detection algorithm of MIMO system, the most commonly used algorithms are Zero Forcing (ZF) and Minimum Mean Square Error (MMSE).

3.1 Zero Forcing Algorithm

According to the above established models, the amplify signal of relay is $\boldsymbol{X}$, the channel matrix is $\boldsymbol{H}_i^{\mathrm{H}}$, the noise of user k is $\boldsymbol{n}_k$, and the noise of receiving terminal is $\boldsymbol{r}_k$. In (2), there are two parts of the noise. The total noise can be expressed as

$$\boldsymbol{z} = \boldsymbol{n}_k + \boldsymbol{r}_k. \tag{3}$$

We can get the reduction matrix $\boldsymbol{W}$ by obtaining the matrix of the received signal and the known channel information. The purpose of signal detection is to obtain the approximate original signal $\widetilde{\boldsymbol{X}}$ through the reduction matrix $\boldsymbol{W}$. In the ZF algorithm [6–8], the reduction matrix is

$$\boldsymbol{W}_{\mathrm{ZF}} = \left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}, \tag{4}$$

in which, $\tilde{\boldsymbol{x}}_{\mathrm{ZF}}$ is the reduction signal obtained by the ZF algorithm, we can deduce the $\tilde{\boldsymbol{x}}_{\mathrm{ZF}}$ under the reduction matrix by calculating the reduction matrix and the received signal.

$$\tilde{\boldsymbol{x}}_{\mathrm{ZF}} = \boldsymbol{W}_{\mathrm{ZF}}\tilde{\boldsymbol{y}}_k = \boldsymbol{x} + \left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{z}, \tag{5}$$

in which, the error of the reduction signal is expressed as

$$\tilde{\boldsymbol{z}}_{\mathrm{ZF}} = \left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{z}. \tag{6}$$

Further operation of $\tilde{\boldsymbol{z}}_{\mathrm{ZF}}$ is performed, and its expected value is obtained and its Singular Value Decomposition (SVD) is performed. After simplification, it can be expressed as

$$\left\|\tilde{\boldsymbol{z}}_{\mathrm{ZF}}\right\|^2 = \left\|\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{z}\right\|^2 = \left\|\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{z}\right\|^2, \tag{7}$$

in which, $\boldsymbol{U}$ is unitary matrix, $\boldsymbol{A}$ is diagonal matrix, and $\boldsymbol{V}$ is the eigenvector of $\boldsymbol{H}$ matrix.

According to its properties, it can be derived as

$$E\left\{\left\|\tilde{\boldsymbol{z}}_{\mathrm{ZF}}\right\|^2\right\} = E\left\{\left\|\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{z}\right\|^2\right\} = \left\|\boldsymbol{V}\boldsymbol{A}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{z}\right\|^2 = \sum_{i=1}^{N_{\mathrm{t}}}\frac{\sigma_z^2}{\sigma_i^2}, \tag{8}$$

in which, σ_z^2 is the variance of channel noise, and σ_i^2 is the noise variance of the channel corresponding to the transmit antenna i.

3.2 Minimum Mean Square Error Algorithm

Unlike the ZF algorithm, the core idea of MMSE algorithm is to maximize the signal-to-noise ratio of subsequent detection, so that the variance of the signal reduced from the received signal to the original signal is minimized.

Therefore, the MMSE algorithm adds channel noise σ_z^2 to the design of the reduction matrix [6], and its reduction matrix is:

$$\boldsymbol{W}_{\mathrm{MMSE}} = \left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \sigma_z^2\boldsymbol{I}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}. \tag{9}$$

The MMSE algorithm needs a better and faster estimation of the channel noise in the process of implementation. Therefore, it is necessary to calculate the reduction matrix more quickly and accurately. After counting the noise σ_z^2 and completing the calculation of the reduction matrix, the received signal is restored. The signal restored after the MMSE algorithm is

$$\tilde{\boldsymbol{x}}_{\mathrm{MMSE}} = \boldsymbol{W}_{\mathrm{MMSE}}\tilde{\boldsymbol{y}}_k = \boldsymbol{x} + \left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \sigma_z^2\boldsymbol{I}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{z}. \tag{10}$$

Similarly, the error of the reduced signal is

$$\tilde{\boldsymbol{z}}_{\mathrm{MMSE}} = \left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \sigma_z^2\boldsymbol{I}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{z}. \tag{11}$$

Further operation of $\tilde{\boldsymbol{z}}_{\mathrm{MMSE}}$ is performed, and its expected value is calculated. Its Singular Value Decomposition (SVD) is applied as

$$\left\|\tilde{\boldsymbol{z}}_{\mathrm{MMSE}}\right\|^2 = \left\|\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \sigma_z^2\boldsymbol{I}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{z}\right\|^2 = \left\|\boldsymbol{V}\left(\boldsymbol{\Lambda} + \sigma_z^2\boldsymbol{\Lambda}^{-1}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{z}\right\|^2, \tag{12}$$

in which, $\boldsymbol{U}$ is the unitary matrix, $\boldsymbol{\Lambda}$ is diagonal matrix, and $\boldsymbol{V}$ is the eigenvector of the $\boldsymbol{H}$ matrix [9].

According to its properties, it can be derived as

$$E\left\{\left\|\tilde{\boldsymbol{z}}_{\mathrm{MMSE}}\right\|^2\right\} = E\left\{\left\|\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \sigma_z^2\boldsymbol{I}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{z}\right\|^2\right\} = \sum_{i=1}^{N_{\mathrm{t}}}\frac{\sigma_z^2\sigma_i^2}{\left(\sigma_i^2 + \sigma_z^2\right)^2}, \tag{13}$$

in which, σ_z^2 is the channel noise error, and the σ_i^2 is the noise variance of the channels corresponding to the i transmitting antenna [10].

By contrasting the MMSE Algorithm with the ZF algorithm, the main difference between them is that the influence of noise on the MMSE algorithm is less than that of the ZF algorithm for the same noise. Theoretically, if the σ_i^2 of each channel is far greater than σ_z^2, then the performance of the ZF algorithm and the MMSE algorithm should be similar.

4 Simulation Result

The experiment is based on Rayleigh channel and Rician channel [11]. The MIMO channel capacity [12] is

$$C = \log_2 \det\left(\boldsymbol{I}_{N_R} + \frac{SNR}{N_T}\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right), \tag{14}$$

the simulation results are obtained and shown in Figs. 2 and 3. Figure 2 shows that the channel capacity of MIMO relay system increases with the increase of the number of transmit antennas under Rayleigh channel. The result in Fig. 3 is under the Rician

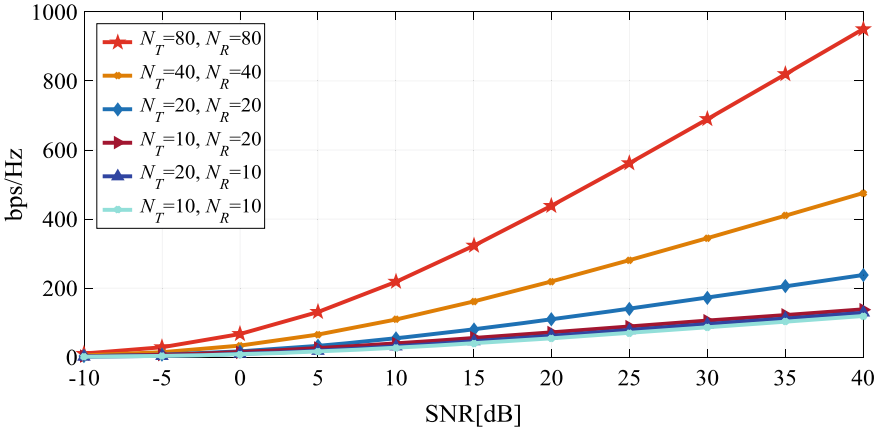


Fig. 2. The channel capacity of MIMO relay system under Rayleigh channel

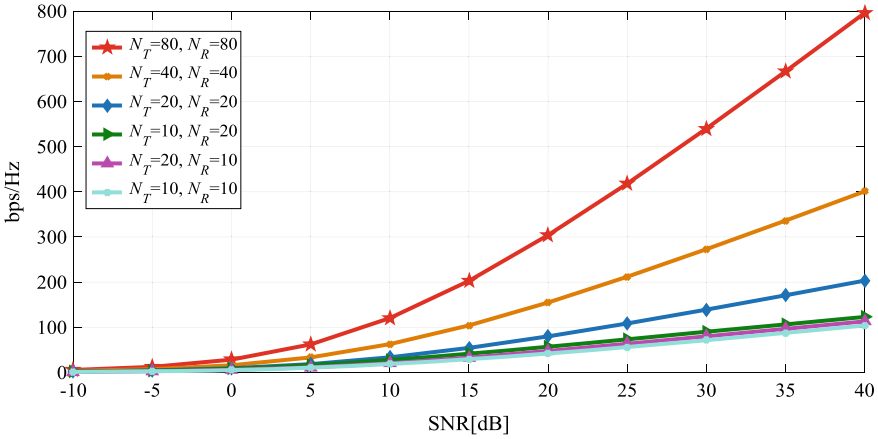


Fig. 3. The channel capacity of MIMO relay system under Rician channel

channel, where the Rician factor is 5 dB. In contrast to Fig. 2, we can clearly find that under the two channels, the multiuser MIMO relay system can increase the number of antennas to enhance the capacity of MIMO relay system.

The number of transmit antennas is 4, and the number of receiving antennas is also 4. Under Rayleigh channel, two different signal detection methods are used to simulate the hardware damage for different degrees, and the output is shown in Fig. 4. It can be clearly seen that with the increase of hardware damage level (ξ_R and ξ_k are both from 0 to 0.005), the bit error rate of the system has increased significantly. When other conditions are the same and MMSE is used for signal detection, the bit error rate of MMSE is lower than that of ZF.

With the same hardware damage degree ($\xi_R = \xi_k = 0.001$), the system is simulated by means of increasing the number of antennas in Rayleigh channel, and the output is shown in Fig. 5. It can be seen that when there is certain degree of hardware damage,

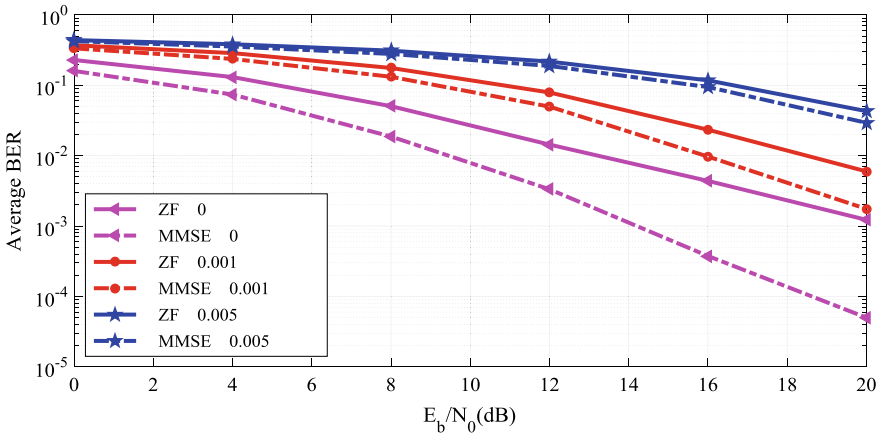


Fig. 4. Average BER in Fig. 1 vs. different degrees of hardware damage

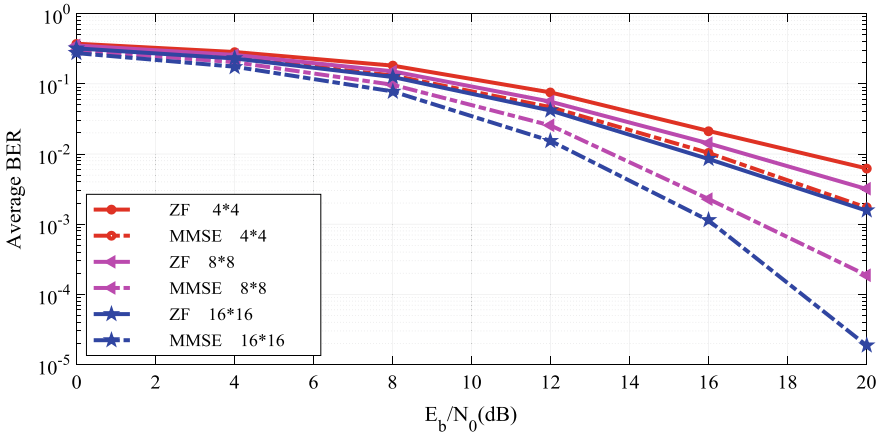


Fig. 5. Average BER in Fig. 1 vs. the number of different antennas

the bit error rate of the system decreases with the increase of the number of antennas. When the number of antennas is the same, the bit error rate is lower when the MMSE signal detection method is used.

5 Conclusion

In this paper, the multiuser MIMO relay system model is analyzed, and a hardware damage model is established for the receiving terminals of the MIMO relay system. An independent zero mean Gaussian noise is added to each antenna, and the variance is proportional to the power of the receiving signal. The simulation results show that the influence of hardware damage has a significant impact on the BER of the system.

In the multiuser MIMO relay system, increasing the number of antennas (include transmitting and receiving antennas) can reduce the bit error rate of the system. In addition, we find that increasing the number of antennas in Rayleigh and Rician channels can significantly improve the capacity of the system. Therefore, in practical engineering applications, people should periodically detect the antennas at MIMO relay stations and replace the antennas with poor performance in time. At the same time, the capacity of the system can be increased by increasing the number of antennas to meet the users' needs.

Acknowledgments. This work was supported by Hainan Provincial Key R. & D. Projects of China (ZDYF2016010 and ZDYF2018012) and the National Natural Science Foundation of China (No. 61661018). Hui Li is the corresponding author.

References

1. Gustavsson, U., Sanchez-Perez, C., Eriksson, T., Athley, F., Durisi, G., Landin, P., et al.: On the impact of hardware impairments on massive MIMO. In: 2014 IEEE Globecom Workshops, GC Wkshps 2014, pp. 294–300 (2014)
2. Sun, X.L., Xu, K., Ma, W.F., Xu, Y.Y., Xia, X.C., Zhang, D.M.: Multi-pair two-way massive MIMO AF full-duplex relaying with imperfect CSI over Ricean fading channels. IEEE Access **4**, 4933–4945 (2016)
3. Zhang, Z.S., Chai, X.M., Long, K.P., Vasilakos, A.V., Hanzo, L.: Full duplex techniques for 5G networks: self-interference cancellation, protocol design, and relay selection. IEEE Commun. Mag. **53**(5), 128–137 (2015)
4. Riihonen, T., Werner, S., Wichman, R.: Hybrid full-duplex/half-duplex relaying with transmit power adaptation. IEEE Trans. Wirel. Commun. **10**(9), 3074–3085 (2011)
5. Singal, A., Kedia, D., Jaglan, N., Gupta, S.D.: Performance analysis of MIMO-OFDM system with transceiver hardware impairments. In: 4th IEEE International Conference on Signal Processing, Computing and Control, ISPCC 2017, pp. 102–105 (2018)
6. Yong, S.C., Jaekwon, K., Won, Y.Y., Chung, G.K.: MIMO-OFDM Wireless Communications with MATLAB. Wiley Publishing (2010)
7. Younas, T., Li, J.D., Arshad, J., Tulu, M.M.: Performance analysis of improved ZF algorithm for massive MIMO in uplink. Electron. Lett. **53**(23), 1554–1556 (2017)

8. Yin, Z.L., Mao, X.P., Zhang, Q.Y., Zhang, N.T.: Fast recursive algorithm for implementation of MIMO ZF-SIC detection. J. Commun. **33**(7), 67–74 (2012)
9. Ghanbarisabagh, M.: Minimizing performance loss of MMSE algorithm for CDMA-based O-MIMO-OFDM. Optik **125**(22), 6801–6804 (2014)
10. Huang, Z.Y., Balanis, C.A.: The MMSE algorithm and mutual coupling for adaptive arrays. IEEE Trans. Antennas Propag. **56**(5), 1292–1296 (2008)
11. Li, X.W., Huang, M.Y., Tian, X.J., Guo, H., Jin, J., Zhang, C.S.: Impact of hardware impairments on large-scale MIMO systems over composite RG fading channels. Int. J. Electron. Commun. (AEÜ) **88**, 134–140 (2018)
12. Rusek, F., Persson, D., Lau, B.K., Larsson, E.G., Marzetta, T.L., Edfors, O., et al.: Scaling up MIMO: opportunities and challenges with very large arrays. IEEE Signal Process. Mag. **30**(1), 40–60 (2013)

Energy-Balanced Routing Protocol with QoS Constraints in Ad Hoc Network

Lu Guo, Peng Li(✉), Jiyu Jin, and Jun Mou

School of Information Science and Engineering, Dalian Polytechnic University, Dalian 116034, China
18742013323@163.com,
{lipeng,jiyu.jin,moujun}@dlpu.edu.cn

Abstract. With the rapid growth of mobile computing and intelligent terminal services, ad hoc network has been developed rapidly in the field of wireless communication and is considered as an important supplementary form of 5G, because of its self-organization and high spatial multiplexing rate. Since the increasing volume of mobile traffic, battery-powered mobile terminals are faced with severe challenges in terms of energy efficiency and QoS. In this paper, an energy-balanced routing protocol with QoS constraints is designed based on AODV protocol. Energy balance is realized by selecting nodes with higher residual energy of nodes to forward packets, and the objective function is established by using bandwidth and path length to realize the QoS constraints. The simulation results show that the improved routing protocol has been effectively improved in terms of the ratio of energy exhausted nodes, end-to-end delay, packet delivery rate, normalized routing overhead, and so on.

Keywords: Ad hoc network · Energy balance · QoS constraints · AODV

1 Introduction

In recent years, ad hoc networks are one of the popular research areas due to low cost to employ the network, self-organization, freedom of location, and no base stations [1–3]. ad hoc network has been developed rapidly in the field of wireless communication and is considered as an important supplementary form of 5G [4] because of its self-organization and high spatial multiplexing rate. As the ad hoc networks are usually powered by batteries with limited power [5], once the battery energy is exhausted, not only the node itself cannot communicate, but also affect the overall network connectivity, quality of service, and lifetime. Moreover, due to the randomness of the moving speed and the mobile mode of each node in the network [6], the network topology may change at any time, and the channel bandwidth and node energy are limited. Therefore, finding a path that satisfies QoS constraints and can make reasonable use of effective resources to ensure the reliable transmission of data is the key to the study.

The original version of this chapter was revised: Author name has been updated. The correction to this chapter can be found at https://doi.org/10.1007/978-981-13-6264-4_132

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 992–999, 2019.
https://doi.org/10.1007/978-981-13-6264-4_117

In terms of energy-balanced, an energy-aware routing protocol with terminal nodes [7] is proposed, which improved the routing efficiency, but enhanced the time of routing reconstruction; An energy-optimized ad hoc network routing protocol [8] introduced the Lorenz index method to improve the energy efficiency of the network, but led to a slight increase in routing overhead; The improved AODV protocol based on energy balance [9] reduced the routing overhead and end-to-end delay, but the packet delivery rate has not been improved.

In terms of QoS constraints, a new ad hoc network routing protocol was proposed in [10] by Othmen S et al., realized high-level QoS by selecting a multiple path that considers the energy constraint and QoS requirements, but its real-time application performance needed to be improved; An AODV extended protocol based on QoS guarantee was proposed in [11], which provided better QoS for network, but it reduced the lifetime of the network; A new QoS-based routing algorithm AODV-QSRP in [12] considered delay, bandwidth, and link quality as route selection standards to support high QoS but increased the overhead.

In order to overcome the above problems, an energy-balanced routing protocol with QoS constraints is designed based on AODV protocol in this paper. The nodes with higher residual energy are selected to forward packets, and the bandwidth and path length are taken as important parameters to realize QoS constraints.

2 The Improvement of AODV Routing Protocol

Due to the classical AODV routing protocol selects the optimal path with the minimum hop count, the residual energy and QoS problems of the node are not considered. Therefore, the concept of node stability is introduced, nodes with higher residual energy are selected to forward packets to improve the lifetime of the node. Then, we use the bandwidth and path length to establish the objective function and select the local optimal solution to realize the QoS constraints as much as possible.

2.1 Node Stability

The stability of a node refers to the contrast between the residual energy of the node and the average residual energy of its neighbors. Among them, the average residual energy of the neighbor nodes is,

$$E_{ave} = \frac{E_r^i + \sum_{k=1}^{n} E_r^k}{n+1} \tag{1}$$

In Eq. (1), E_{ave} is the average residual energy of neighbor nodes, E_r^i is the residual energy of node n_i, n is the number of neighbor nodes of the node n_i, and E_r^k is the residual energy of the $k\text{-}th(k = 1,2,...,n)$ adjacent nodes.

2.2 The Realization of QoS Constraints

For AODV routing protocol, it is not enough to select neighbor nodes based on node stability, so it is necessary to add QoS constraints mechanism. Known that bandwidth [13] is an important factor to realize QoS constraints, and the distance between nodes [14] also affects the quality of service forwarding. Therefore, this paper regards bandwidth and path length as important parameters to realize QoS constraints.

If there are n neighbor nodes around the node n_i, the distance between the k-th ($k = 1,2,...,n$) neighbor node and the node n_i is w_k and the bandwidth is v_k. Then,

$$m_k = \frac{v_k}{v_{\max}} * \frac{w_k}{\sqrt{X^2 + Y^2}} \tag{2}$$

$$\sum_{k=1}^{n} w_k \leq C \tag{3}$$

In Eq. (2), v_{max} denotes the maximum value of the nodes bandwidth, X and Y denote the length and width of the nodes moving areas, and m_k denotes the product of the bandwidth and the normalized distance of the neighboring node k. Equation (3) is the constraint condition of the path length.

Assuming that the source node $A(x_a,y_a)$ and the destination node $B(x_b,y_b)$, so the shortest path length is,

$$C_{\min} = \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2} \tag{4}$$

Therefore, the path length C is equal to the sum of the shortest path length and the path length that is set independently,

$$C = C_{\min} + C_{set} \tag{5}$$

In Eq. (5), the value of C_{set} depends on the actual situation of the network. If the bandwidth of the nodes in the network is generally larger, then try to select a smaller value of $C_{set;}$ Otherwise, you need to select a larger value of $C_{set.}$

3 Route Discovery and Maintenance

Like the classic AODV routing protocol, the improved AODV routing protocol process is divided into two processes: route discovery and route maintenance.

3.1 Route Discovery

When a service is generated in the network, first, the source node identifies the address and sequence number of the destination node, and then finds the route that whether contains the destination node in the routing table of the source node itself. If so, communicates directly, and if not, start the routing discovery process. The specific steps are as follows:

1. First, the source node initiates a RREQ packet, obtains the residual energy of the neighbor node, and then calculates the network average residual energy according to Eq. (1), updates it in the RREQ packet, and then broadcasts the RREQ packet.
2. The source node compares the residual energy of each neighbor node according to the average residual energy of the network in the RREQ packet, and avoids the node with relatively low residual energy.
3. The source node obtains its bandwidth information from the neighbor nodes that is suitable for forwarding RREQ, and calculates the distance from the source node to each neighbor node. Based on Eq. (2), the m_k value of each neighbor node is calculated and compared, and the neighbor node with the largest m_k value continues to forward and broadcast the RREQ packet until the destination node is reached.
4. When the RREQ packet arrives at the destination node, the total length of the path is calculated. If the path length C is satisfied, the route discovery succeeds and the RREP message is sent by the destination node. If the RREQ packet is not satisfied, it is discarded and the route discovery process is restarted according to the AODV protocol.
5. After receiving the RREP message, the source node forwards the data according to the next hop node's receiving condition and until the destination node receives the data successfully.

3.2 Route Maintenance

The improved AODV routing protocol calculates the average residual energy of the network and the residual energy of the neighbor nodes through the Hello message, and determines the average level of the residual energy in the network. Then "abandons" the neighbor nodes with low stability, and effectively avoids the neighbor nodes with insufficient residual energy. To a certain extent, it avoids the situation of circuit breakage caused by energy exhaustion at the center node.

4 Simulation Results

In order to verify the network performance of the improved routing protocol, the performance of the ratio of energy exhausted nodes, end-to-end delay, packet delivery rate, and the normalized routing overhead of the two routing protocols is simulated.

There are 50 nodes in a network, which move over a square 1000 × 1000 m flat space, each node has a total energy of 50 J, *and* each packet size is 512-bytes, 20 CBR flows are generated randomly at a rate of 100 packets per second. The simulation time is 600 s and the Max. speed is 30 m/s.

The following analysis of four kinds of performance under different simulation time and Max. speed. Among them, the moving speed of the node is 20 m/s for different simulation time, the simulation time is 600 s for different node movement speeds. And we can see that the improved routing protocol has been effectively improved in terms of the ratio of energy exhausted nodes, end-to-end delay, packet delivery rate, normalized routing overhead, and so on.

4.1 The Ratio of Energy Exhausted Nodes

Figure 1(a) shows that the improved protocol selects nodes with higher average residual energy to forward packets, the ratio of energy exhausted nodes in the improved protocol is relatively reduced. Figure 1(b) because the improved protocol selects nodes with higher stability to forward packets, the ratio of energy exhausted nodes is relatively small.

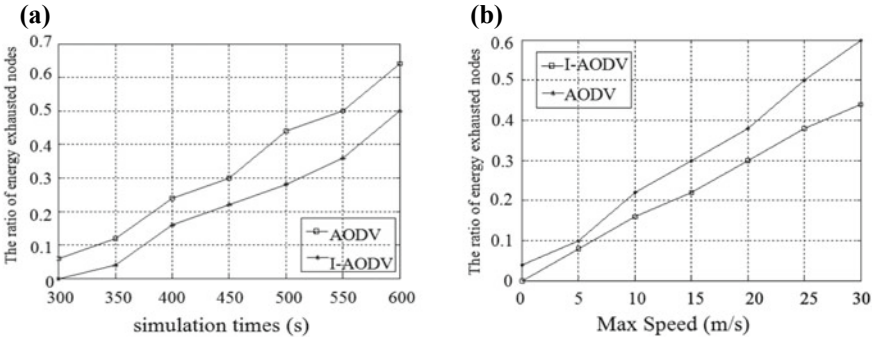


Fig. 1. The ratio of energy exhausted nodes under different simulation time Vs Max. speed. (a) simulation time Vs the ratio of energy exhausted nodes, (b) Max. speed Vs the ratio of energy exhausted nodes

4.2 The Average End-to-End Delay

Figure 2(a) shows that the improved protocol uses nodes with larger bandwidth and residual energy to forward packets, and limits the total length of the path, which reduces the transmission time of packets and then reduces the delay. Figure 2(b) shows that the improved protocol uses the bandwidth and path length to establish the objective function and selects the appropriate nodes to forward data.

4.3 The Packet Delivery Rate

Figure 3(a) shows that after 300 s, the improved routing protocol can increase the lifetime of the node because of the selection of the nodes with high residual energy, which reduces situation of circuit breakage caused by energy exhaustion at the nodes. Figure 3(b) because the improved AODV protocol can select nodes with higher stability to forward packets and enhance the stability of the link.

4.4 The Normalized Routing Overhead

Figure 4(a) shows that the improved protocol selects nodes with higher residual energy to forward RREQ messages, which improves the stability of the links and reduces the number of packets maintained by the route. Figure 4(b) because the improved protocol

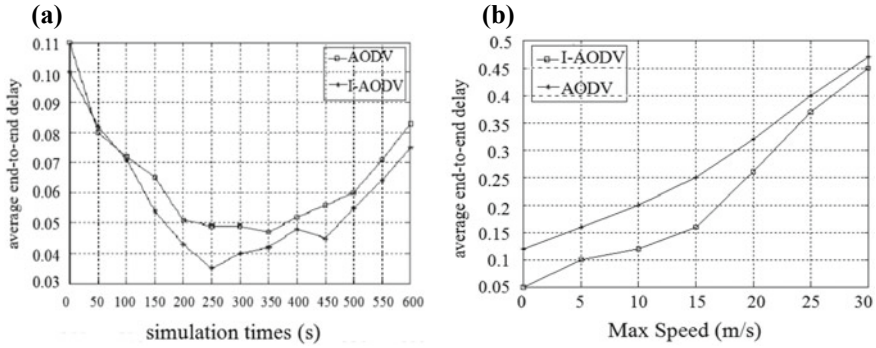


Fig. 2. The end-to-end delay under different simulation time Vs Max. speed. (a) simulation time Vs end-to-end delay, (b) Max. speed Vs end-to-end delay

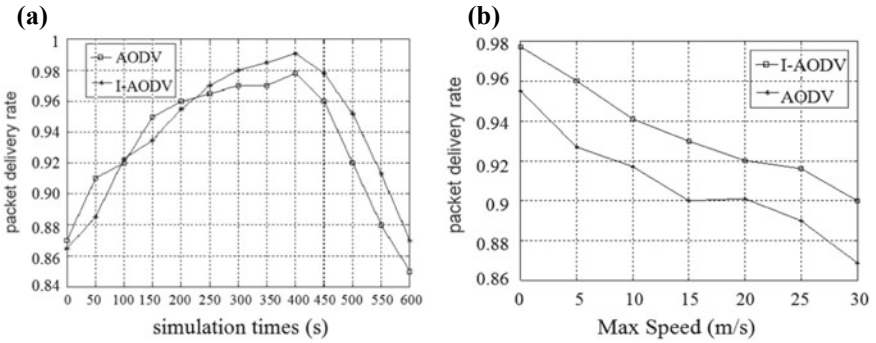


Fig. 3. The packet delivery rate under different simulation time Vs node movement speeds. (a) simulation time Vs packet delivery rate, (b) Max. speed Vs packet delivery rate

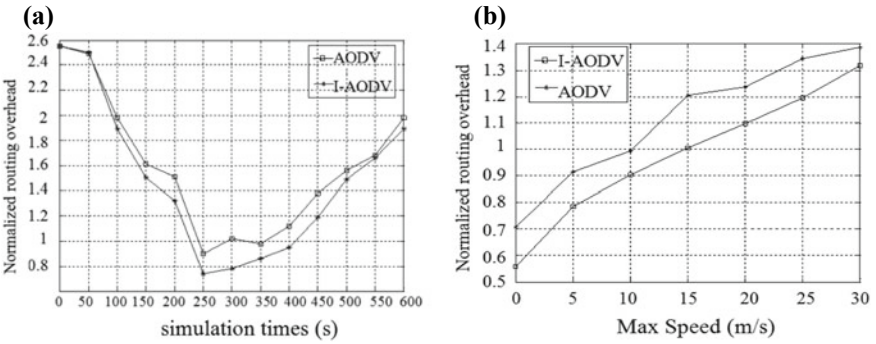


Fig. 4. The normalized overhead under different simulation time Vs node movement speeds. (a) simulation time Vs normalized overhead, (b) Max. speed Vs normalized overhead

selects nodes with higher residual energy to forward packets, reduces the ratio of energy exhausted nodes, increases the stability of the network, and reduces the process of routing discovery and routing maintenance.

5 Conclusion

Aiming at the shortage of AODV routing protocol, this paper puts forward two improvements from node stability, bandwidth and path length, which are mainly embodied in energy-balanced and QoS-constrained. It can be seen from the simulation that, the improved AODV routing protocol improves the network performance in four aspects of the ratio of energy exhausted nodes, average end-to-end delay, packet delivery rate, and the normalized routing overhead.

Although the improvement of AODV routing protocol has achieved some results through simulation, but there are still some shortcomings. For example, there are many parameters to realize the QoS constraints. Whether it is possible to achieve better results by establishing the objective function in addition to other parameters outside the bandwidth and the path length, whether the improved routing protocol is better than other improved routing protocols should be further discussed and studied.

Acknowledgements. This paper is supported by Natural Science Foundation of Liaoning, China (2015020031), and Basic Scientific Research Projects of Colleges and Universities of Liaoning, China (2017J046).

References

1. Aldrin, S.K., Perkins, C.: An energy efficient multipath routing protocol based on signal strength for mobile ad-hoc network. ARPN J. Eng. Appl. Sci. **11**, 7370–7375 (2016)
2. Zhan, T., Wang, L.: Research on mobile database model of ad hoc network based on multi-parameter weighted clustering. C E Ca, vol. 10, pp. 11–22 (2017)
3. Allahham, A.: Performance evaluation of DSR, AODV, and OLSR routing protocols in MANET. In: The International Conference on Computer Science & Computational Mathematics, pp. 1–8 (2016)
4. Farris, I., Orsino, A., Militano, L., Iera, A., Araniti, G., Farris, I., et al.: Federated iot services leveraging 5 g technologies at the edge. Ad Hoc Netw. **68**, 58–69 (2018)
5. Yu-jun, L., Yi-bo, W., Meng, C.: Ad-hoc improved hybrid routing protocol based on energy equalization. J. China Acad. Electron. Inf. Technol. **12**, 486–492 (2017)
6. Yong-hong, K., Chun-na, G., Jian, C.: Energy optimization and load balancing QoS on-demand routing protocol. J. Appl. Sci. **30**, 25–30 (2012)
7. Zheng, S., Zhang, P., Zhang, Q.Y.: A routing protocol based on energy aware in ad hoc networks. J. Commun. **33**, 9–16 (2012)
8. Hong, L., Weijun, X., Qitao, W.: Research on energy-balanced ad hoc network routing protocols. Comput. Meas. Control. **24**, 204–207 (2016)
9. Fanshuo, K., Yang, N.: Improvement and simulation analysis of AODV protocol based on energy balance. Internet Things Technol. **7**, 41–45 (2017)
10. Akbani, R., Korkmaz, et al.: Mobile ad-hoc networks security. Recent. Adv. Comput. Sci. Inf. Eng. **127**, 659–666 (2015)

11. Kumar, N., Zeadally, S., et al.: Qos-aware hierarchical web caching scheme for online video streaming applications in internet-based vehicular ad hoc networks. IEEE Trans. Ind. Electron. **62**, 7892–7900 (2015)
12. Di Filippo, M., Lucci, L., Marabissi, D., et al.: Design of a smart antenna for mobile ad hoc network applications. Int. J. Antennas Propag. 1–7 (2014)
13. Sharma, L., Lal, C., Sharma, D.P., et al.: Enhancing QoS for multimedia services using mobility-aware bandwidth estimation algorithm in MANETs. In: International Conference on Optical & Wireless Technologies, pp. 655–666 (2017)
14. Debbarma, J., Debbarma, M.K., Roy, S., et al.: An efficient energy protocol for power conservation in mobile ad-hoc networks. In: International Symposium on Computational and Business Intelligence, pp. 239–242 (2014)

Research on UHF RFID Anti-collision Algorithm Based on Collision Slots

Chong Wang and Weiping Jing(✉)

School of Electronics and Information,
Nantong University, Nantong 226019, China
304709776@qq.com, 13906294039@163.com

Abstract. The problem of anti-collision is an important research topic in UHF RFID system. In order to further improve the recognition efficiency of RFID system, a new anti-collision algorithm based on collision slots elimination is proposed. The algorithm sends a new command through the reader to filter the tags of the collision slot one by one, making the tags in the collision slot further identified, that can avoid tags in the collision slot to enter the next frame recognition. Because the tags of collision slot are identified, the number of frame slots is reduced, and the coordination time in the recognition process is saved. Finally, the theoretical analysis and simulation results show that the new algorithm, based on the ISO 18000-6C protocol, reduces the recognition time by about 25% and improves the recognition rate by about 27% compared with the 18000-6C protocol.

Keywords: UHF RFID · Anti-collision · Collided slot · ISO 18000-6C · Recognition rate · Recognition time

1 Introduction

Radio Frequency Identification (RFID) technology is one of the core technologies in the concept of Internet of Things [1]. RFID systems have common frequencies such as low frequency, high frequency, ultrahigh frequency, and microwave. When the low-frequency and high-frequency readers read the tags, the metal and the human body have less influence on them, but the reading distance is relatively short. Compared to other frequency band radio frequency identification systems, UHF RFID systems operating at 860 MHz–960 MHz have the advantages of long reading distance, fast reading speed of multiple tags, and strong anti-interference ability, which are currently hot spots for the development of RFID technology in the world [2].

The tag collision problem is undoubtedly a thorny problem encountered by the readers in all frequency bands. The collision problem of the tag is a key factor that limits the reading speed and affects the accuracy of the reader, so how to solve this problem is very important. This paper focuses on the anti-collision algorithm for UHF bands. Tag anti-collision algorithm is mainly divided into non-deterministic anti-collision algorithm and deterministic anti-collision algorithm [3]. The anti-collision algorithm defined by the ISO 18000-6C protocol is a non-deterministic anti-collision

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 1000–1008, 2019.
https://doi.org/10.1007/978-981-13-6264-4_118

algorithm—a random time slot algorithm based on a random number generator [4]. Reference [2] is the prediction of frame slots. The algorithm proposed in reference [5] is to optimize the Q value to shorten the number of slots. Most anti-collision algorithms based on the ISO 18000-6C protocol do not further deal with collision slots. This not only results in missing reading of the tags but also affects the reading speed of the reader. Based on the 18000-6C protocol, this paper introduces a new algorithm to further deal with collision time slots, so as to achieve the purpose of reducing slots and improving the recognition rate of the tags.

2 Random Slot Algorithm Based on Random

This algorithm is mentioned in the ISO 18000-6C protocol. The main flow is shown in Fig. 1. The algorithm contains commands such as Query, Query Rep, Query Adjust, etc. [5]. The main parameter of the algorithm is parameter Q of the slot counter. Q is generated by the reader's random number generator, and the value is an integer in the range of 0–15.

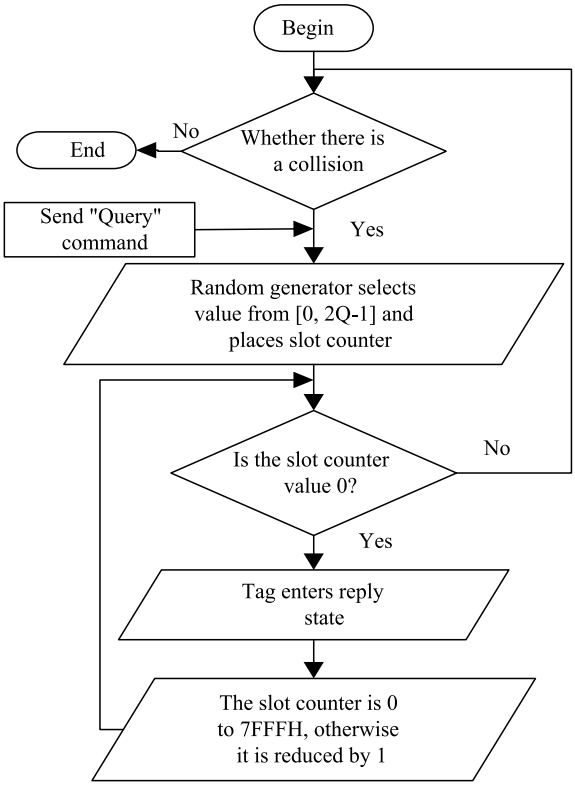


Fig. 1. 18000-6C protocol algorithm

First, when the reader detects tag collision, it sends the "Query" command that carries the parameter Q. When the tags which in the antenna coverage area receive a command, the random number generator of tag generates a number from the range of 0 to 2Q-1 and places it in the slot counter of the tag. The tag whose slot counter is 0 enters the reply state, otherwise it enters the arbitration state and waits for the next slot detection. When the tag receives the "Query Rep" command, the value of the slot counter is decremented by one, and if the slot counter is 0, it will be changed to 7FFFh [6, 7]. It is to prevent them from continuing to participate in the reply. The reply will not be resumed until the tag generates a new random number. When the tag receives the "Query Adjust" command, the reader adjusts the size of the Q value and sends it again to the tag. When only one tag enters the reply state, the tag's random number generator generates a 16-bit random number, denoted as RN16 [8, 9], which is backscattered to the reader and the reader acknowledges the tag with an ACK containing the same RN16. The tag which is confirmed immediately enters the confirmation state and continues to backscatter its PC, EPC, and CRC-16. If many tags enter the reply state, the reader detects the collision by detecting the waveform. If the reader can resolve the RN16 which is sent by the tag, the tag will be acknowledged by the reader and the tag will backscatter its PC, EPC, and CRC-16 [10, 11]. Tags that are not recognized by the reader receive the wrong RN16. They will not send their PC, EPC, and CRC-16 and return to arbitration state.

3 Improved New Algorithm

3.1 New Algorithm Description

Nowadays, most studies are aimed at improving the Q value. However, after the tags have been filtered by slots, there is still collision which means many tags are in the response state, and thus they enter the collision slot. It has no effect to identify the tags of the collision slot with the existing methods. The tags can only wait until the detection of the next frame, which undoubtedly increases the recognition time. The reference [8] proposes to use binary search anti-collision method to deal with the collision slots problem, but this algorithm ignores the problem that binary search anti-collision method is based on Manchester encoding, and the encoding of the data transmission between the tag and the reader is based on FM0 (Bi-Phase Space) or Miller encoding [12]. This method does not comply with the 18000-6C protocol standard.

Therefore, this paper proposes a new algorithm for further processing of collision slots. The specific process is shown in Fig. 2.

The reader detects collision when the slot counter of many tags is 0 and the random number RN16 is backscattered. Then the reader enters the above algorithm.

For example, the slot counter of four tags is 0, and the random number RN16 which they generate is as follows:

Tag1 ……1 0 1 0
Tag2 ……1 1 0 0
Tag3 ……1 0 0 1
Tag4 ……1 0 1 1

When the reader detects the signal and the signal collides, it sends the Filter command to the tag, which carries the parameter f. The initial value of f is 0 and the size is between 0 and 15. The tags of this slot receive the command to judge whether the f-th bit of its own RN16 is 0, and if the tag whose RN16[f] is 0 sends its own RN16 to the reader. Because the initial value of f is 0, when the reader sends the Filter command for the first time, the condition that the RN16[0] is 0 is met. Only tag 1 and tag 2 respond to the reader and backscatter their RN16 [13, 14]. At this point, the reader will detect the collision again, and the reader will continue to send the Filter command and adjust the value of f such that f = f+1. The condition that RN16 [1] is 0 is involved in the response, then only the tag 2 responds to the reader. The reader continues to send the Filter command, and f restores the initial value of 0. Only tag 1 satisfies the condition that RN[0] is 0, so tag 1 responds to the reader, and f restores the initial value of 0. The reader detects that the collision still exists and continues to send the Filter command. At this point, there is no tag that satisfies the condition that RN[0] is 0. But the reader can still detect the existence of the collision. Then, the reader adjusts the value of f, so that f = f + 1, only tag 3' RN [1] is 0, then tag 3 responds the reader. Finally, when the tag 4 sends the RN16, no tag collision occurs and the reader can detect the tag 4 successfully. So far, all four tags have been detected. When all tags of the slot have been detected, the reader sends the Query_Rep command to start the identification of the next slot.

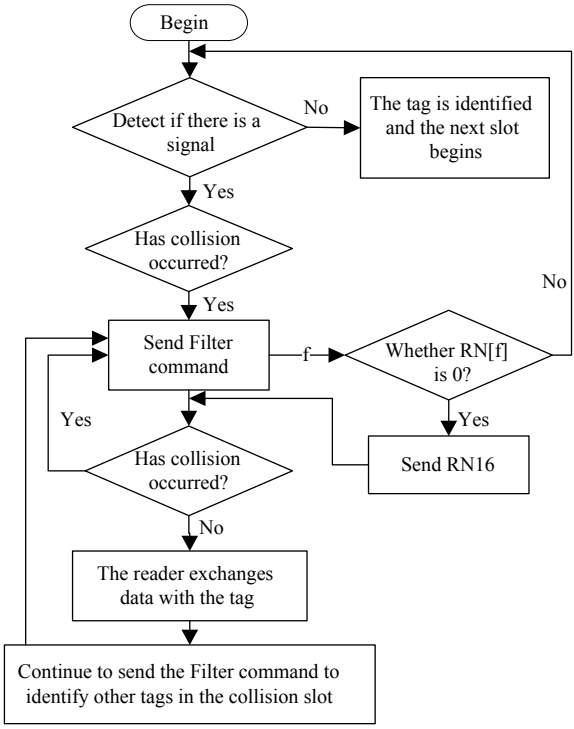


Fig. 2. New algorithm flow

The advantage of the new algorithm is that it solves the problem that the slots collision in the 18000-6C protocol cannot be handled. In the specified time T [15], the reader recognizes the tags within the collision slots as much as possible, thereby reducing time that reader detect tags, improving the work efficiency of the reader and improving the tag recognition rate.

3.2 Algorithm Analysis

If the reader reads 100 tags, Q is 4 which is the default value, and collision slots are processed by the new algorithm. The following table is obtained by analyzing reader can read only one tag or two tags successfully at each collision slot.

Table 1. Algorithm for identifying a tag in a collision slot

The value of Q	Total number of slots (L)	The number of tags started (m)	Number of successful slots E(1)	Number of collision slots E(r)	Number of idle slots E (0)	Unidentified tags
4	16	100	0	16	0	84
5	32	84	6	24	2	54
6	64	54	23	13	27	18
5	32	18	10	3	18	5
4	16	5	4	1	12	0

Table 2. Algorithm for Identifying Two Tags in A Collision Slot

The value of Q	Total number of slots (L)	The number of tags started (m)	Number of successful slots E(1)	Number of collision slots E(r)	Number of idle slots E (0)	Unidentified tags
4	16	100	0	16	0	68
5	32	68	8	20	4	48
6	64	48	23	11	30	3
5	32	3	3	0	29	0

From Tables 1 and 2, it can be seen that if a tag is identified in each collision slot, 160 slots are required to read 100 tags. When two tags are identified in each collision slot, the reader reading 100 tags requires 144 slots and the number of slots is reduced by 54.5% and 71.6%, respectively. It can be seen from the comparison of the results that if the tags can be identified within the collision slot, the total number of slots can be greatly reduced, thereby reducing the tag identification time and improving the system's work efficiency.

4 Simulation and Result Analysis

This paper uses Matlab software to model the RFID system to analyze the changing relationship between the number of frame slots, recognition time, and recognition rate when the new algorithm and the 18000-6C protocol algorithm changing with the number of tags. In each simulation experiment, the tag's EPC is represented by a 12-bit hexadecimal number, and the number of tags is 10, 25, 50, 80, 100, 120, 140, 160, 180, 200. To prevent from random values appearing during the simulation, each point is simulated 50 times and then averaged.

The relationship between the number of frame slots and the number of tags is shown in Fig. 3.

As can be seen from Fig. 3, when the number of tags is 100, the number of time slots of the new algorithm is reduced by about 77% compared with the 18000-6C algorithm, and the total number of frame time slots is reduced by about 39%.

The identification time changing with the number of tags is as shown in Fig. 4.

As can be seen from Fig. 4, when the number of tags is more than 50, the recognition time of the new algorithm is reduced by approximately 25% compared with the identification time of the 18000-6C protocol algorithm.

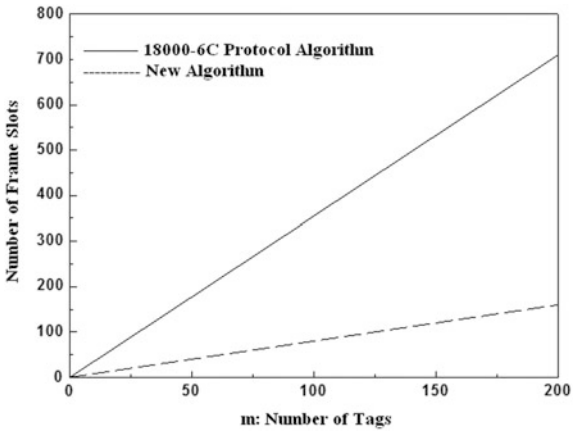


Fig. 3. The relationship between the number of frame slots and the number of tags

The comparison between the new algorithm and the 18000-6C algorithm is analyzed from the tag recognition rate.

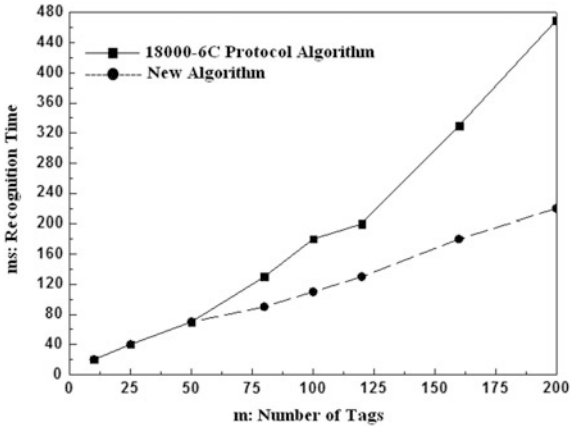


Fig. 4. Relationship between the number of tags and identification time

Reference [12] shows that when two tags collide, the number of slots is:

$$E(2) = \frac{m-1}{2} \cdot (1 - \frac{1}{L})^{m-2} \tag{1}$$

When the time slot in which two tags collide is recognized by the reader, the recognition rate of the tag is:

$$S = \frac{E(1) + E(2)}{L} \tag{2}$$

From the formula (2), it can be seen that the recognition rate is higher than the recognition rate of the collision time slot tag that is not recognized by E(2)/L. If the collision time slots of r(r > 2) tags can be recognized by the reader, the recognition rate will be greatly improved.

The comparison between the recognition rate of the new algorithm and the recognition rate of the 18000-6C algorithm is shown in Fig. 5.

From Fig. 5, it can be seen that the recognition rate of the new algorithm and the 18000-6C protocol algorithm are gradually reduced with the increase of the number of tags. When the number of tags approaches 200, the recognition rate tends to be stable. When the number of tags is 180, the recognition rate of the new algorithm is about 0.6379, while the recognition rate of the 18000-6C protocol algorithm is about 0.3689. The recognition rate of the new algorithm is about 27% higher than that of the 18000-6C protocol algorithm.

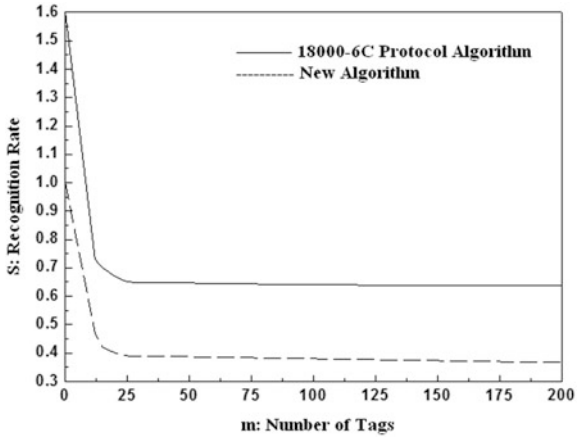


Fig. 5. Relationship between the number of tags and recognition rate

5 Conclusions

This paper proposes a new algorithm based on the 18000-6C protocol to further deal with collision time slots. The reader sends the filter command to filter the tags in the collision time slot one by one and identifies the tags in the collision time slot one by one, which solves the problem that tags cannot be identified in the collision time slot. The algorithm can effectively shorten the time slot and improve the recognition rate of the tag. The simulation results show that the new algorithm is shorter than the 18000-6C algorithm and the recognition rate is improved. In the next step, the improved method of anti-collision algorithm in different applications will be further studied to solve the multi-tags collision problem in reality.

Acknowledgements. The fund project of this paper is: Special fund project for transformation of scientific and technological achievements in Jiangsu. The number of the project is BA2017065.

References

1. Vales-Alonso, J., Francisco, J., Alcaraz, J.J.: Analytical computation of the mean number of tag identifications during a time interval in FSA. IEEE Commun. Lett. **18**(11), 1923–1926 (2014)
2. Yuan-Cheng, L., Ling-Yen, H., Bor-Shen, L.: Optimal slot assignment for binary tracking tree protocol in RFID tag identification. IEEE Trans. Netw. **23**(1), 255–268 (2015)
3. Tian, Y., Kang, H.: Research of improved ALOHA anti-collision algorithm in RFID system. In: Communications, Signal Processing, and Systems (2018)
4. Barlettt, L., Borgonovo, F., Filippini, I.: Asymptotic analysis of schoute's estimate for dynamic frame Aloha. In: International Conference on Software, Telecommunications and Computer Hetworks, pp. 113–118. Splits IEEE (2015)

5. Lijuan, Z., Wei, X., Xiaohu, T.: An adaptive anticollmion protocol for large-scale RFID tag identification. IEEE Wirel. Commun. Lett. **3**(6), 601–604 (2014)
6. Yuan-Cheng, L., Ling-Yen, H., Bor-Shem, L.: Optimal slot assignment for binary tracking tree protocol in RFID tag identification. IEEE Trans. Netw. **23**(1), 255–268 (2015)
7. EPC Global, Radio-Frequency Identity Protocols :Class 1 Generation! UHF RFID Protocolfor Communications at 860 MHz-960 MHz: Versionl. 2. 0 [EB/OL]. [2016-11-21]. https://www.gsl.org/sites/default/files/docs/epc/uhfclg2_1_0_9-standard-20050126.pdf
8. Wu Y., Yao S.: A modified slotted random anti-collision algorithm for the EPC global UHF Class-1 Generation-2 Standard. In: Proceedings of IEEE International Conference on Computational Science & Engineering. Washington D. C., pp. 397–400. IEEE Press, USA (2014)
9. Chen, W.T.: Optimal frame length analysis and an efficient anti-collision algorithm with early adjustment of frame length for RFID systems. IEEE Trans. Veh. Technol. **65**(5), 3342–3348 (2016)
10. Zhang, Y., Yang, F., Wang, Q., et al.: An anti-collision algorithm for RFID-based robots based on dynamic grouping binary trees. Comput. Electr. Eng. (2017)
11. Jian, S., Sheng, Z., Xie, L., et al.: Idle-slots elimination based binary splitting anti-collision algorithm for RFID. IEEE Commun. Lett. **20**(12), 2394–2397 (2016)
12. Su, J., Sheng, Z., Xie, L.: A collision-tolerant based anti-collision algorithm for large scale RFID system. IEEE Commun. Lett. **99**, 1–1 (2017)
13. Zhang, Y.J., Cui, Y.: EDMC: an enhanced distributed multi-channel anti-collision algorithm for RFID reader system. In: International Conference on Materials Science. AIP Publishing LLC, pp. 821–832 (2017)
14. Xin, W., Jia, Q.X., Xin, G., et al.: Research on grouping n-ray tracking tree RFID anti-collision algorithm. Acta Electron. Sinica (2016)
15. Lin, Y.H., Liang, C.K.: A highly efficient predetection-based anticollision mechanism for radio-frequency identification. J. Sens. Actuator Netw. **7**(1), 13 (2018)

Research on Unmanned Aerial Vehicle Based Ad Hoc Network Incorporating Speed and Energy Awareness

Shaowei Sun, Bingcai Chen(✉), Haifang Li, Mei Nian(✉), and Weimin Pan

School of Computer Science and Technology, Xinjiang Normal University, Urumqi, China
cbc9@qq.com

Abstract. In order to provide a reliable and stable UAV ad hoc network, we propose a routing strategy that combines the existing ad hoc routing protocol and optimized link protocol (OLSR) and implements the UAV as a carrier. A speed-aware and energy-aware routing protocol is adopted, and the mobility of nodes and the load balancing of nodes are comprehensively considered in the clustering algorithm. This paper focuses on the design, implementation, and evaluation of the UAV's ad hoc network routing strategy. By simulating with OMNeT ++, we demonstrate the superiority of this strategy.

Keywords: UAV · OLSR · Cluster algorithm · Speed-aware · Energy-aware

1 Introduction

Nowadays, compared with a single Unmanned Aerial Vehicle (UAV) system, a multi-UAV system can accomplish a task by working together, making it more efficient and less costly [1]. This requires us to provide a reliable network. However, there are still many problems in building a network of UAVs to provide a stable and reliable network and the Ad hoc network has become a multiple UAV system structure. Due to the particularity of UAV ad hoc network, the consistency must also be considered as more nodes are needed for high mobility, high dynamic changing topology, link of intermittent, energy saving and changeful link quality. In order to improve the performance of UAV ad hoc network, this paper will fully consider the speed and energy issued in the UAV ad hoc network to ease the problem of the UAV ad hoc network topology in the high dynamic change and the limitation of the UAV energy.

2 Difficulties

In this section, we will introduce the difficulties of UAV ad hoc network from three aspects, which are high-speed movement, node energy, and clustering.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 1009–1013, 2019.
https://doi.org/10.1007/978-981-13-6264-4_119

2.1 High Dynamic Changes in Network Topology and Intermittent Nature of Links Caused by High-Speed Movement of Nodes

The high-speed movement of nodes is the most significant difference between UAV ad hoc networks and traditional ad hoc networks. This high-speed movement will bring high dynamic changes to the topology which will have a serious impact on the performance of the network and the performance of the protocol [2]. The communication failure of the UAV platform and the instability of the line-of-sight communication link will also cause link interruption and topology update. And in the UAV ad hoc network, due to the random movement of the UAV, the mutual interference between the wireless channels, the comprehensive factors such as the terrain, the network topology formed by the wireless channel between the mobile terminals will change at any time, and the manner of change and speed are unpredictable, which makes routing extremely important. This requires us to fully consider the speed issue in the research.

2.2 Node Energy Limited

In various wireless ad hoc networks particularly the UAV ad hoc network, there is a prominent problem that the available energy of the node is limited.

In the UAV ad hoc network, the UAV nodes in the network generally need to conduct many tasks. After completing these tasks, the load of the UAV is not too high, that is, the battery carried by the UAV node is not too large. In general, the UAV node consumes energy mainly in two aspects: the first aspect is that the task will consume energy; the second aspect is the energy consumed by sending and receiving data, including the control command packets and routing information packets. This requires us to achieve fair energy consumption per node, so that the life of the entire network can be greatly improved.

2.3 The Impact of Cluster Size Selection on the Network Quality

The clustering algorithm logically separates the Ad hoc network according to a certain strategy, and separated regions are connected to each other. The main problems involved include the selection of the cluster head and the selection of the size of the cluster. When the network topology changes, the algorithm needs to effectively maintain the entire network.

The cluster size selection affects the implementation of the data link layer algorithm. After the clustering is determined, the topology of the network is fixed, and the data link layer begins to implement the MAC layer algorithm. For the contention MAC layer protocol, the average waiting time T of each node to access a channel is related to the cluster size. The parameter affected by the cluster size is the communication density Pe. The relationship between the two is proportional. The formula is as follows:

$$T \propto Pe \tag{1}$$

When the number of each cluster increases, it is obvious that the Pe of the cluster, the average waiting time T of the node connected to the data channel also increases

while the communication delay performance will decrease. However, if the size of each cluster is reduced, the total number of clusters will increase, and the number of backbone members will increase as well, which will degrade the performance of the backbone network

3 Related Work

In this section, we focus on the above problems and introduce two strategies we have adopted.

3.1 Integrating Energy-Aware and Speed-Aware Concepts into the OLSR Protocol

The core idea of OLSR is the MPR mechanism [3]. Each node selects a MPR node to broadcast the packet, so that the size of the link state update message packet is greatly reduced, and the network overhead can be reduced. However, the high dynamic movement of the UAV nodes will cause the high dynamic changes of the network topology. Some UAV nodes selected as MPR nodes cannot be selected as MPR nodes because of their mobility, it will re-trigger the MPR nodes elections. Therefore, the entire network will reelect MPR nodes, and network reestablishment of routing will lead to an increase in end-to-end delay.

In this paper, we use the multipath OLSR routing protocol and backup path of OLSR routing protocol considering the idea of balanced load, [4] and Dijkstra algorithm in routing protocol (MBCR) combined with minimum battery metrics to calculate the second optimal path of data packet transmission, and save it. This can greatly reduce the data transmission delay. Among them, the minimum battery consumption routing (MBCR) is a cost function which is inversely proportional to the residual energy of nodes to distribute the energy consumption between nodes fairly [5]. The equation is shown as follows:

$$F(ki) = 1/ki, \tag{2}$$

where ki is the remaining energy of the node.

3.2 Clustering Algorithm Considering Cluster Head Load and Cluster Stability

First of all, there already exist some common clustering algorithms, such as the minimum ID clustering algorithm and the link clustering algorithm. The algorithm has fast convergence rate and low maintenance overhead to consider load balancing, but the cluster head node consumes energy a lot, which will reduce the lifetime of the entire network and bring about the problem of excess cluster heads.

The above clustering algorithms do not consider the fairness factors, so we incorporate speed-aware and energy-aware concepts into the algorithm. In the process of cluster head election, the combination of weighted method is used to conduct

election. The expected value of nodes with high mobility as cluster heads is small, and that of nodes with low energy is also small. Therefore, we weighted the combination of speed and energy of nodes to calculate a priority value. The larger the priority value is, the more likely the cluster head is to be elected. The equation is shown as follows:

$$Weight = a^{*}Mobility + b^{*}Power + c^{*}Energy - left \tag{3}$$

where Mobility represents the moving speed of the node, Power represents the transmission power, and Energy-left represents the remaining energy. a, b, c are set according to the specific environment of the network. At the same time, we need to set a threshold. In the algorithm, the node not only has the node degree but also needs to maintain another variable election degree. This variable indicates the node degree when the node is elected as the cluster head. The node selected as the cluster head calculates the difference between the current two variables. If the absolute value of the difference exceeds the previously specified threshold, the cluster head qualification is canceled, otherwise it is still the cluster head. It can be a good solution to the situation where a node is elected as a cluster head for a long time.

Finally, we will introduce the parameters used in the simulation module. The UAV ad hoc network simulation parameters are shown as follows:

Parameters	Values
Network topology range	100 km * 100 km
Transmission distance	50 km * 50 km
Number of network nodes	10
Movement speed	100 m/s
Transmission rate	10 Mbps
Packet length	12000 bit
Packet interval	Poisson distribution
Simulation time	600 s

4 Results and Conclusion

We first use the proactive routing protocol OLSR which can meet the requirements of time delay for tasks to integrate speed-aware and energy-aware concepts into the OLSR protocol which will greatly improve the UAV ad hoc network performance. Then, we utilize the clustering routing algorithm considering both node energy and node speed, thus the cluster head election mechanism can be optimized, which reduces the load of the cluster head and improves the performance of the whole network. The UAV ad hoc network is simulated in the OMNeT ++ system [6]. The experimental research shows the feasibility of the proposed routing strategy, which makes some improvement in the end-to-end delay, throughput rate, and packet loss rate in the UAV ad hoc network.

Acknowledgement. This work was supported by NSFC (61771089) and Tendering Research Project of the key disciplines of the Xinjiang Normal University (17SDKD1201).

References

1. Gupta, L., Jain, R., Vaszkun, G.: Survey of important issues in UAV communication networks. IEEE Commun. Surv. Tutor. **18**(2), 1123–1152 (2016)
2. Yanmaz, E., Yanmaz, E., Muzaffar, R.: Survey on unmanned aerial vehicle networks for civil applications: a communications viewpoint. IEEE Press (2016)
3. Clausen, T., Jacquet, P.: Optimized link state routing protocol(OLSR). RFC 3626 (2003)
4. Le, M., Park, J.S., Gerla, M.: UAV assisted disruption tolerant routing. In: Military Communications Conference, 2006. Milcom, IEEE, pp. 1–5 (2006)
5. Ma, X., Kacimi, R., Dhaou, R.: Adaptive hybrid MAC protocols for UAV-assisted mobile sensor networks. In: IEEE Consumer Communications & Networking Conference. IEEE (2018)
6. Rtner, E, Vom Lehn, H, Wehrle, K.: A performance comparison of recent network simulators. In: IEEE International Conference on Communications. IEEE, pp. 1–5 (2009)

Spectrum Optimization

A Frequency-Domain Composite Preamble to Integrate Sensing and Communication

Xiaokun Zheng(✉), Ting Jiang, and Wenling Xue

Key Labs of Universal Wireless Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China
1053072637@qq.com

Abstract. In order to integrate sensing and communication signal, a method of replacing WiFi preamble with composite preamble is proposed. The signal can perform timing sync and fine frequency offset (FO) estimation, etc., meanwhile there is a better recognition resolution. Without extra bandwidth occupied, ① A direct synthesis is formed by overlying short time Chirp signal on original in time domain; ② a complex frequency domain synthesis is formed by using IFFT of signal which has a similar spectrum with the direct synthesis. FDTD simulation for 802.11a shows that recognition resolution is improved. On the other hand, anti-fading of direct synthesis estimation is slightly affected. Direct synthesis has some engineering value. Simulation also shows that anti-fading of complex synthesis is unaffected at all, and the complex synthesis can be used for more rigorous scenarios.

Keywords: Composite preamble · Integration · Sensing and communication · Channel estimation · Target recognition

1 Introduction

Integrated sensing and communication is a hotspot issue in communication terminal. It is important to health care, military [1], transport [2], and industry monitor, border patrol, etc. It is important to select suitable integration signal, which meets the function of communication and sensing simultaneously.

Using widely distributed WiFi or RFID signal for recognition is an attractive technology. In [3, 4], WiFi signal is used for human motion and fall detection, [4] reported 87% fall detection accuracy when the magnitude of CSI was used. [5, 6] discussed multistatic passive radar experiment using WiMAX and RFID, their sensing resolution are generally limited. How to improve the recognition accuracy and resolution of WiFi is a state-of-the-art issue.

The work was supported by national natural science foundation of China 61671075.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 1017–1026, 2019.
https://doi.org/10.1007/978-981-13-6264-4_120

One way is to select different physical characteristics or improve the recognition algorithm. Physical characteristics include extracting signal power, delay, kurtosis, and CSI multipath information [3], MIMO perception antenna forming multi-dimensional features [7], multiple WiFi continuous data symbols make FFT together [4], etc. Recognition algorithm mainly includes SVM, deep learning and neural network classification algorithm. Such as the method based on Fourth-order Cumulants [8], or using dynamic time warping to improve SVM recognition rate [9], etc.

The other way is waveform design, by appropriately modifying the WiFi signal, its communication is maintained meanwhile its perception capability is improved. The most important way is to mix radar Chirp signal. [2] uses time division mode and sets up a fixed slot to chirp signal; [7, 10] use frequency division mode and allocate subcarriers to sensing signal; [11] improves perception resolution by lengthening and rebuilding CP(cyclic prefix); In [12, 13], communication sequence and sensing signal are simultaneously transmitted by code division. The above methods are all signal multiplex modes, sensing signal always occupy resource in time, frequency, or in code domain. Thus, the data frame and transmission rate are affected somewhat.

Integrated waveform design is studied in this paper. To obtain a higher communication performance than normal multiplex integration, we replace original WiFi preamble with a combined long preamble, meanwhile it does not occupy an extra bandwidth or time slot. The combined preamble can perform channel estimation, such as frequency offset (FO) estimation, equalization, etc., meanwhile enhance the perception resolution of original preamble. Data section of WiFi changes constantly, it is difficult to use data section to sensing [4]. So it is a good choice to use preamble, and a composite signal by using Chirp and original preamble is proposed in this paper. Theoretical analysis and simulation based on 802.11a show that the sensing resolution can be improved. On the other hand, anti-fading performance of direct synthesis channel estimation is slightly affected. For example in deep fading, the probability of FO estimation error within 5 kHz decreases a bit. But error within 10 kHz is closed to 100%, which meets the requirements of 802.11a [14]. The direct synthesis method has some engineering value. In addition, an IFFT of a signal which has a similar spectrum with direct synthesis can be used. Its fading resistance is as good as original preamble, it can be considered for more rigorous scenarios.

The rest of this paper is organized as follows. In Sect. 2, the composite method and analytic expression are given, and then correlation and frequency character are drawn out. In Sect. 3 first, simulation shows that FO estimation of direct time synthesis is slightly affected in fading channel. Secondly, simulation shows that by using IFFT of signal which has a similar spectrum, the anti-fading is as good as original preamble. In Sect. 4, FDTD (finite difference time domain) simulation shows that recognition resolution is improved by using synthesis preamble. In Sect. 5, paper concludes that the direct time synthetic has engineering value. And since the anti-fading of complex frequency synthesis is as good as original preamble, it can be used for more rigorous scenarios.

2 Scheme of the Composite Preamble

2.1 Composite Preamble

Frequency domain of 802.11a long preamble is constant amplitude, original preamble L_a is IFFT of L = {00 000 1, 1, −1, −1, 1, 1, −1,1, −1, 1 …, −1, 1, 1, 1, 1, 000000}, or say L_a = IFFT(L). And according to table G.6 of 802.11a standard [15], real part in time is La_r = [−0.156, 0.012, 0.092, −0.092, −0.003…], imag part is La_i = [0, −0.098, 0.106, −0.115, −0.054, 0.074 …].

The common radar chirp signal is $s(t) = U(t)e^{-j2\pi t(f_0 + \frac{1}{2}f_{slope}\cdot t)}$. Composite preamble method (1) of this article, or say direct time synthesis can be expressed as:

$$P_r(t) = \begin{cases} L_a(t) & 0 \le t \le T_L - T_{chrp} \\ L_a(t) + s(t) & T_L - T_{chrp} < t \le T_L \end{cases}, \tag{1}$$

where T_L is the length of original La, and one cycle T_L= 3.2us. T_{chrp} is the length of short time chirp, and $s(t)$ does not occupy extra bandwidth. Figure 1b shows the spectrum of the composite signal in Eqs. 1 and 2.

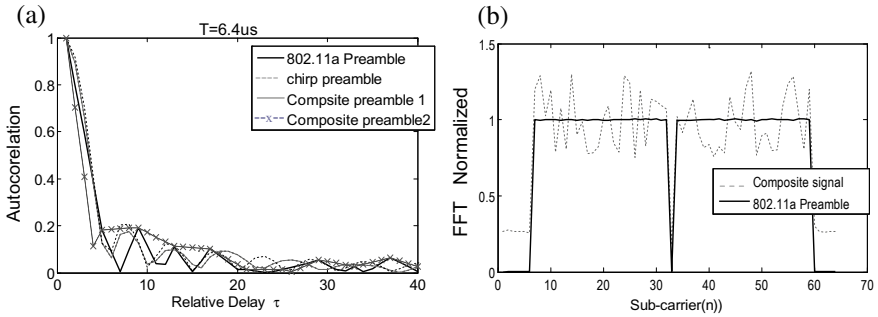


Fig. 1. Time and Frequency domain character (a) Autocorrelation; (b) Frequency domain character

T_{chrp} of direct synthesis cannot be too long, otherwise the influence on anti-fading is a bit great. And if T_{chrp} is too short, the improvement of perception resolution is not obvious. We set T_{chrp}= 0.5us, the recognition rate improvement is shown in Fig. 5, meanwhile channel estimation is easy to meet requirements.

According to the frequency feature of direct time synthesis, it can be considered to use IFFT of the signal which has similar spectrum structure and replace original constant amplitude spectrum. F-domain composite preamble or say complex synthesis (2) can be written as:

$$P_r'(t) = IFFT\left[L + abs(FFT(k \cdot s(t)))\right] \tag{2}$$

2.2 Autocorrelation Character and F-Domain Character

This section we examine the time and frequency domain features of the synthetic signal.

$$R(\tau) = \frac{1}{T} \int_0^T f^*(t) f(t+\tau)\, dt \tag{3}$$

Equation 3 is the autocorrelation function. Where duration $T = 2T_L = 6.4$ us, delay $\tau = nT_s$. And according to USRP data sheet (universal software radio peripheral) [20], we set front sample rate $f_s = 80$ MHz then $T_s = 1/80$ M(a)

Figure 1a shows the time domain property. Autocorrelation of the composite signal, 802.11a preamble and chirp signal are included in Fig. 1a, it can be seen that the three are almost the same. As described in Sect. 3.1, chirp signal along is not suitable as a preamble, because envelope signal is greatly affected by deep fading.

Figure 1b is frequency-domain property. Preamble of 802.11a is constant amplitude frequency domain, and the synthesized signal fluctuates based on the constant amplitude, simulation shows that synthetic signal has higher perception resolution (see Fig. 4). Though there is a small fluctuation on basis of original constant amplitude, frequency domain equalization (FDE) estimation can still be performed. The difference between FFT of received composite and transmit composite can be used as FDE compensation value. FDE is not the focus of this article.

3 Channel Estimation Simulation

In this section, we will simulate channel estimation with synthetic preamble. (1) We first replace the original preamble directly with chirp, estimation performance is poor in fading channels; (2) Then direct synthesis is used for frequency estimation, it is slightly affected in deep fading. However, it can satisfy the FO error requirement of 802.11a [14]. Direct time synthesis has some engineering value; (3) IFFT of signal which has similar spectrum is used, its anti-fading is as good as original WiFi. It can be considered for more rigorous scenario

3.1 Chirp Preamble

Since radar chirp signal itself has good correlation feature (Fig. 1a), we first think of replacing the original preamble directly with chirp.

In Fig. 2, the dark solid and dotted line correspond to the 802.11a preamble show that the estimation is completely unaffected by fading. The light curve is the chirp preamble, and the light dotted line corresponds to a deeper fading. It can be seen that estimation of chirp is greatly affected by fading, so pure chirp is not suitable as a preamble. The reason is that the envelope signal is greatly affected by fading.

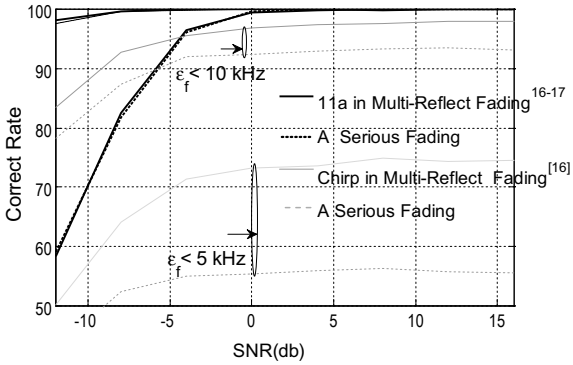


Fig. 2. Frequency offset estimation of the Chirp preamble

3.2 Channel Estimation of Time Synthesis

In this section, we simulate fine time sync and FO estimation by using time domain synthesis, then Fig. 3 is drawn.

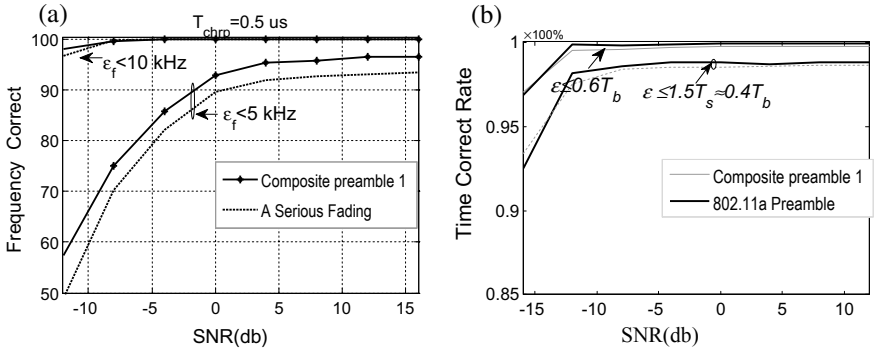


Fig. 3. Fine time sync and FO estimation of direct time synthesis (a) Frequency offset estimation; (b) Fine time sync

Figure 3a is a simulation result of fine FO estimation for direct synthesis. Marked line corresponds to general multi-reflect fading [16, 17], Mean attenuation of each path is set as [0 −3 −10 −20 −30] dB. And attenuation of more serious fading is set as [0 −1.8 −9 −10 −15] dB. It shows that probability of error within 5 kHz is reduced by about 2% when serious fading, but error within 10 k is almost the same as original, is closed to 100%. That satisfies the 802.11a requirement that the FO error should be within 10 kHz when 54 Mbps [14].

Figure 3b is the result of fine timing. Front sample rate is set to 80 MHz, thus sampling period T_s = 1/80 M. Symbol period of preamble T_b = 1/20 M, then each symbol is sampled 4 times T_s = 1/4 T_b. When sync judgment is at the exact correct point, sync error $|\varepsilon| < 1/8\ T_b$. When the fine sync decision is allowed to miss at the next 1 sampling points, $|\varepsilon| < 3/8\ T_b \approx 0.4\ T_b$. In Fig. 3(a) the light line is composite signal, in general multipath fading performance decreased slightly, but probability of $\varepsilon < 0.4\ T_b$ is close to 98% like original preamble, $\varepsilon < 0.6\ T_b$ is closed to 100%.

3.3 Channel Estimation of Frequency Synthesis

In this section, we mainly simulate FO estimation by using frequency domain synthesis to illustrate anti-fading characteristic when channel estimation is implemented, then Fig. 4 is drawn.

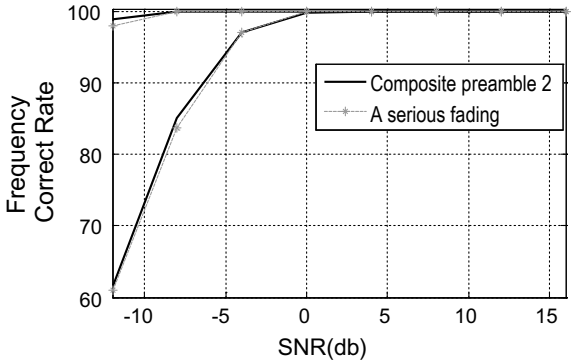


Fig. 4. FO estimation of complex frequency synthesis

Figure 4 is the FO estimation of complex frequency synthesis. It shows that the performance has hardly any deterioration in serious fading; the fading resistance is as good as original WiFi. Frequency synthesis can be considered for more rigorous scenarios.

3.4 Assessment of the Communication Capability

According to Table 90 in 802.11a [15] (”Allowed RCE versus data rate”), we obtain the allowed rate under different RCE and Fig. 5 is drawn.

Figure 5 shows the assessment of the communication capability compared to time division multiplex integration [2], when they are applied to the same OFDM modulation based on 802.11a. We can see that the composition method is superior.

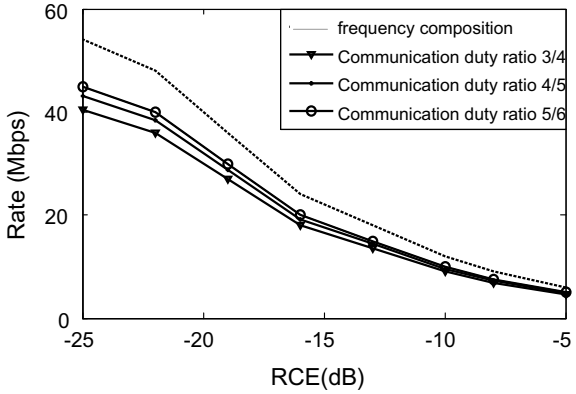


Fig. 5. Assessment of the communication capability vs. time multiplex integration based on 802.11a

4 High-Resolution Target Recognition

4.1 Simulation Settings

This paper focuses on the WiFi preamble. The scene was chosen as indoor environment (see Fig. 6a). The synthesized signal is modulated to 5.2 GHz of the 802.11a, interval of signal transmit and receive is 1.6 m, and the signal is received for recognition when the intermediate objects are 25 cm and 21 cm spheres respectively

Closed scenario could be viewed as a waveguide; FDTD simulation provides more accuracy than general multi-path mode, so we use FDTD method [18, 19] for accurate computation. And after FDTD calculation we add white noise. Table 1 lists main parameters of the scenario.

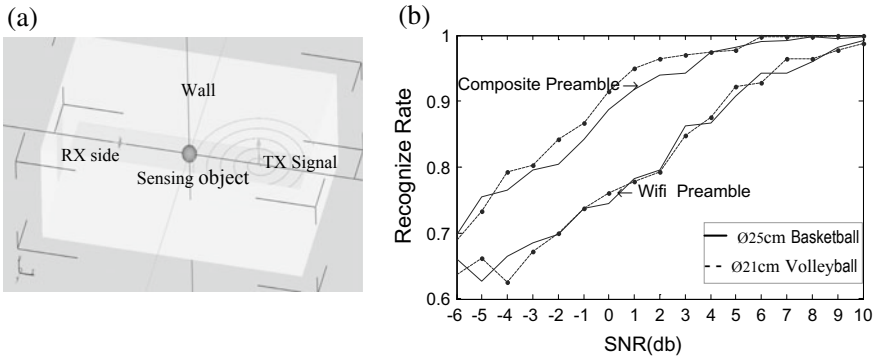


Fig. 6. Recognize scenario and Result (a) Recognition scenario; (b) Recognition result

Table 1. Recognize scenario variables

Variable	Value	Meaning
eps	3	Object ε_r
x_length	260	Length (mm)
y_length	260	Width (mm)
Thickness	30	Thick (mm)
x_space	2	Lab Height (m)
y_space	1.5	Lab width (m)
z_space	2.9	Lab length (m)
L_trans_re	1.6	Distance RX & TX
Location	0.5	Object in the middle
L_antena	12	Length of antenna (*Ndle*)

Table 2. Recognize parameter setting

Basketball	Φ25 cm	Porosity	0.97
Football	Φ22 cm	Material	Rubber
Volleyball	Φ21 cm	Permittivity	$\varepsilon_r = 7.4$
Training sample	600	Test samples	1400
x,y,z Boundary: Absorb boundary, PML = 6			
Time domain Features: Excess delay, std deviation Maximum value, peak value, etc.			

On the basis of normalization of preamble, time domain features of composite signal were extracted, including excess delay, energy, rms delay, maximum value, standard deviation, peak value, etc. SVM classification svmtrain and svmpredict function in Matlab2014 were used in both experiments. Radial Basis Kernel was set, parameter c,g were set to 1 and 10, respectively. 600 training samples and 1400 test samples of each ball were selected. Table 2 shows the main recognize parameter.

4.2 Experimental Results

Figure 6b is the classification results of the same shape and different size objects. The solid line corresponds to the classification accuracy of 25 cm sphere. The dashed line is the classification accuracy of 21 cm sphere. It shows that recognition rate of subtle difference objects is effectively improved, whether in high SNR or low SNR condition.

5 Conclusions

This paper proposed a composite preamble method to integrate sensing and communication. The study focusing on 802.11a shows that channel estimation of direct synthesis can satisfy the 11a requirement in relative deep fading, and recognition resolution has been improved.

Combining the spectrum features of direct synthesis, IFFT of signal which has a similar spectrum can be considered, and replace original preamble which has a constant amplitude spectrum. Simulation shows it has the same channel estimation capability as original WiFi. Its high-resolution recognition performance can be further tested.

Acknowledgments. This work was supported by National Natural Science Foundation of China (NSFC) (No. 61671075) and Major Program of National Natural Science Foundation of China (No. 61631003).

References

1. Yang, Q.: Progress in directional spatially aware wireless network technology of US Navy. J. Mil. Commun. Technol. **36**(1), 59–65 (2015). http://www.cnki.net/
2. Liang, H.A.N., Ke, W.U.: 24-GHz integrated radio and radar system capable of time-agile Wireless communication and sensing. IEEE Trans. Microw. Theory Tech. **60**(3), 619–631 (2012)
3. Adib, Fadel, Katabi, Dina: See through walls with WiFi! ACM SIGCOMM **43**(4), 75–86 (2013)
4. Jiang, H., CaiBrown, C.: Smart home based on WiFi sensing: a survey. IEEE Access **6**, 13317–13325 (2018)
5. Webster, T., Higgins, T.: Passive multistatic radar experiment using WiMAX signals of opportunity. Part 2: Multistatic velocity backprojection. IET Radar Sonar Navig. **10**(2), 248–255 (2016)
6. Zou, Y., Xiao, J.: GRfid: a device-free gesture recognition system using COTS RFID device. IEEE Trans. Mob. Comput. (2016). https://doi.org/10.1109/tmc.2016.2549518
7. Sit, Y.L.: MIMO OFDM radar with communication and interference cancellation Features. In: IEEE Radar Conference, Cincinnati, pp. 19–23 (2014)
8. Zhong, Y., Zhou, Z., Jiang, T.: A novel gesture recognition method by Wi-Fi communication signal based on fourth-order cumulants. In: IEEE ICC Workshop on Radar and Sonar Networks, pp. 10567–10571 (2015)
9. Zhou, G., Ting, J.: A new method of dynamic gesture recognition using Wi-Fi signals based on DWT and SVM improved by DTW. In: IEEE Global Conference on Signal and Information Processing (Global SIP), pp. 12.14–12.18 (2015)
10. Mishra, A.K., Inggs, M.: FOPEN capabilities of commensal radars based on whitespace communication systems. In: IEEE Electronics, Computing and Communication Technologies, Bangalore, pp. 1–5 (2014)
11. Zhang, T., Xia, X.: OFDM synthetic aperture radar imaging with sufficient cyclic prefix. IEEE Trans. Geosci. Remote Sens. **53**(1), 394–404 (2015)
12. Takase, H., Shinriki, M.: A dual-use radar and communication system with complete complementary codes. In: IEEE 15th International, Radar Symposium (IRS) (2014)

13. Jamil, M., Zepernick, H.-J.: On integrated radar and communication systems using oppermann sequences. In: IEEE International Conference on Military Communication. San Diego, America, pp. 1–6 (2008)
14. Li, S., Li X.: Research on 802.11a frequency deviation measurement. In: Application of Electronic Technique, pp. 48–50 (2005). ISSN: 0258-7998
15. IEEE Std 802.11a Part 11: Wireless LAN Medium Access Control MAC and Physical Layer (PHY) specifications (1999)
16. Rayleigh fading channel simulation (2013). https://wenku.baidu.com/view/14be526ae45c3b3567ec8bf7
17. Ling-hong, Kong: Simulation of OFDM system in multipath Rayleigh fading channels. Comput. Simul. **25**(7), 105–108 (2008)
18. Maloney, G., Smith, G.S.: Accurate computation of the radiation from simple antennas using the finite-difference time-domain method. In: Digest on Antennas and Propagation Society International Symposium (1989)
19. Ohtani, T.: A stability improvement technique using PML condition for the three-dimensional nonuniform mesh nonstandard FDTD method. IEEE Trans. Magnet. 1569–1572 (2013)
20. Firas Abbas Hamza (2008) The USRP under 1.5X Magnifying Lens Jun. http://gnuradio.org/trac/attachment

Interference Suppression Under Allocation of Intensive Spectrum Resource with Time-Varying Filter Application

Liu Xuejiao, Sha Xuejun(✉), Chen Zhongyao, and Li Jing

School of Electrical and Information Engineering,
Harbin Institute of Technology, Harbin, China
Shaxuejun@hit.edu.cn

Abstract. In this paper, we discuss the application of time-varying filter in intensive spectrum resource allocation, which is constructed basing energy distribution of the pulse shaping signal. First, we analyze the system model and discuss the existence of interference when in its practical application. Then, we contrast the application of time-varying filter and traditional time-invariant bandpass filter with fixed width, concluding the benefits of time-varying filter in interference rejection. And considering the interference caused by practical channel with multipath delay and Doppler frequency shift, we proposed joint filter combined channel equalizer filter with time-varying filter, to improve spectrum efficiency and suppress interference in 2-D (*t, f*) plane effectively.

Keywords: Time-varying filter · Interference suppression · Spectrum allocation · Fading channel · Energy distribution

1 Introduction

With the development of mobile communication, the research of 5G technology facing 2020 is in full swing, which promotes torrent of links especially due to IOT. Therefore, it is urgent to further improve the system capacity and plan rationally allocation of spectrum resources [1]. The allocation of spectrum resources in wireless mobile communication is carried out on planning and scheduling the limited resources (spectrum and power), so as to make the system a bigger capacity and better performance, and ensure the coverage and quality of service (QoS) requirements for users. Wireless resource allocation is usually combined with other technologies, like design of scheduling scheme, power allocation/control, interference control, optimization crossing layer, etc. To improve spectrum efficiency, the orthogonality of resource has to be ensured in time, frequency or space domain, for instance TDMA, FDMA, and SDMA technology. In addition, adaptive coded modulation, multi-antenna, and multi-carrier transmission technology can also improve the spectral efficiency.

On the other hand, in the field of signal processing, more and more attentions have been paid to the application of time-varying filter (TVF) in recent years, especially suppressing the interference in time-frequency domain. Some scholar has studied the time-frequency distribution(TFD) characteristics of SC-FDMA signals, then designed

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 1027–1035, 2019.
https://doi.org/10.1007/978-981-13-6264-4_121

pass region according to the TFD of reception signals by Choi-Williams distribution to suppress interference from adjacent cells in the LTE uplink system, which has obtained better Signal to Interference plus Noise Ratio (SINR) performance under additive white Gaussian noise (AWGN) channel [2]. However, when the noise and interference power are relatively large, it is difficult to design the filter pass region.

Existing researches for narrowband interference suppression technology depending on DFT theory have been relatively mature. The development of modern signal processing technology has brought a series of new theories specific to non-stationary signals: short-time Fourier transform (STFT), Gabor transform, wavelet transform, Radon Wigner transform, fractional Fourier transform, etc. These theories have led analysis, estimation, and detection of non-stationary signal hot in this field. The research content at present about interference suppression technology using time-frequency filter is focused on the following three aspects: interference rejection schemes including time-frequency filter based on STFT distribution, adaptive filter, signal integration, and projection filter based on TFD, etc.; studies about time-frequency transform for interference suppression; application researches of interference suppression algorithm based on analysis tools above in various communication systems [3]. In actual communication system, the interference is variable, resulted in that corresponding suppression algorithm and scheme remain to be researched.

This paper is focused on the interference suppression based on time-frequency analysis and TVF under the intensive spectrum resource allocation scenario, based on energy distribution in (t, f) plane, which mainly targets the adjacent frequency interference, narrowband interference and the ISI/ICI caused by fading channel. First, we introduce the allocation method and analyze the interference in this model, then bring in the suppression method by TVF method based on TFD. Finally, we proposed the suppression method as joint channel equalizer with TVF for ISI/ICI influenced by fading channel.

2 Theoretical Basis

2.1 System Model

Traditional resource allocation methods including orthogonal resource sharing and non-orthogonal resource sharing, in which signal's own time-varying characteristics in these methods were ignored. Here, we introduce the effective allocation method, which is based on the energy distribution of pulse shaping signal in the time-frequency domain.

By time-frequency analysis we can find the concentrate and the loose area of signals in the (t, f) plane. According to the spectrum characteristic, resource allocation can been done in the (t, f) domain by sensing the energy slots in (t, f) spectrum between adjacent frequency to improve the spectrum efficiency as shown in Fig. 1.

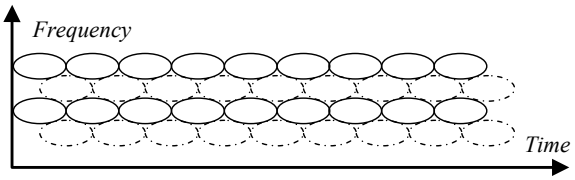


Fig. 1. The time-frequency resource allocation system model

In this system, we need to choose the signal with high energy concentration in (t, f) plane. The document has proved Gaussian pulse has the best time-frequency aggregation characteristics, close to the best energy resolution under the given resolving time [4]. The mathematic expression of Gaussian pulse in time and frequency domain is

$$f(t) = (1/\pi\alpha)^{1/4}\exp\left(-t^2/2\alpha\right) \tag{1}$$

$$F(f) = (4\pi\alpha)^{1/4}\exp\left(-2\alpha\pi^2 f^2\right) \tag{2}$$

The discretization Gaussian signal can be expressed as

$$f(n) = (1/\pi\alpha)^{1/4}\exp\left[-(nT_s)^2/2\alpha\right], \tag{3}$$

where α is the parameter to determine the width of Gaussian pulse. Practically, the pulse is usually truncated for pulse forming, and the aggregation degree is measured by the product of 3 dB bandwidth and symbol duration as BT value. When code {1, 0, 1, 1} is transmitted shaping with BT = 0.4, the time-frequency aggregation is shown in Fig. 2.

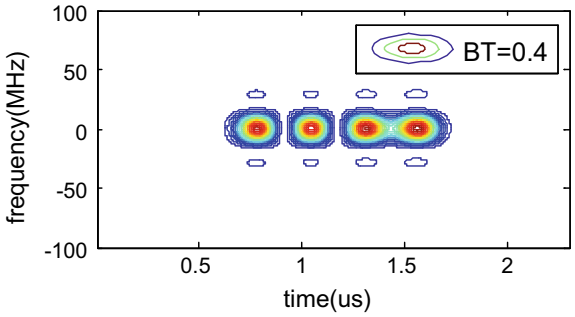


Fig. 2. The time-frequency aggregation degree

We find the time-frequency blocks of symbols in elliptic arrangement and the adjacent symbols are separable. The Gaussian signals have maximum energy in the center of symbol and porosity area at border. When adjacent symbols reverse sign,

porosity areas are more obvious, because phase inversion cause frequency component richer, and low energy area exist within the effective bandwidth.

2.2 Interference Analysis

Under the allocation system of dense spectrum resources above, there are the following interferences. When the channel is the ideal AWGN channel, the received signals by purposed receiver contain interference from the two adjacent users. The simplified model is shown in 0, where Interference from two adjacent users are shown in blue while random Narrow Band Interference (NBI) within the same frequency band may exist in the system, like in yellow. Traditional linear time-invariant (LTI) bandpass filter fixed width is shown in green in Fig. 3, which collects much interference energy.

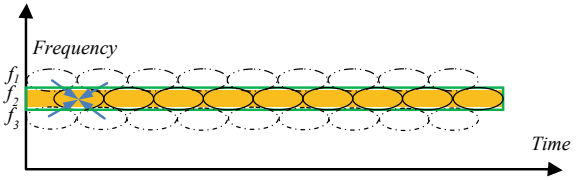


Fig. 3. The interference in allocation system of dense spectrum resources

When in fading channel, the receiver signals contain ISI caused by multipath delay and ICI within the passband caused by Doppler frequency shift [5]. The fading channel model is defined by the parameters including the multipath spread τ, power delay profile $P(\tau)$, maximum Doppler frequency shift f_{dmax}, etc. The delay spread σ_τ is defined as secondary moment of $P(\tau)$. If $\sigma_\tau \ll T_b$, the channel is flat fading channel, otherwise frequency selective fading channel. If the Doppler spread B_D is larger than baseband width B_s, the channel is slow fading channel, otherwise fast fading channel. The fading channel is usually described in tap delay line (TDL) model as shown in (Fig. 4).

Here, we adopt the Jakes model in the simulation to describe the flat fading channel. The received signal at purpose receiver is expressed as

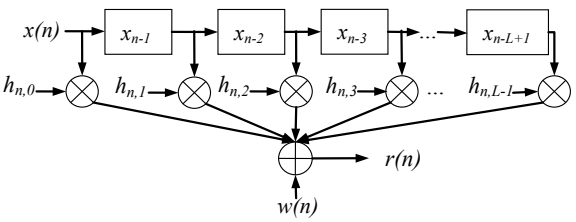


Fig. 4. The TDL model of time-varying multipath channel

$$r(n) = h(n) * x(n) + w(n) = h(n) * [s(n) + \sum i(n)] + w(n), \tag{4}$$

where *h(n)* is channel response, *i(n)* is the adjacent interference, *w(n)* is AWGN noise.

2.3 Time-Frequency Analysis and Time-Varying Filter Theory

The interference rejection algorithm with TVF is new processing technology in transform domain, by converting receiver signals mixed with interferences to another domain using time-frequency analysis tools. In the new transform domain, by analyzing the energy spectrum of corresponding signals, according to signal energy distribution to set pass region, the TVF suppresses most of the interference energy.

The time-frequency analysis maps the 1-D time domain signal into 2-D (*t, f*) plane by utilizing analysis window and shifting it to reflect the localized characteristics of signal spectrum presenting with high energy concentration [6]. In this paper, we utilize the linear time-frequency joint analysis method of Gabor transform, a special case of STFT with Gaussian window [7, 8]. The principle is to give a Gaussian window $g_w(t)$ long for w to capture a time-slice $x_n(t)$ of signal $x(t)$, and obtain the TFD $Ga_{[x(t)]}(t, f)$ by analyzing the localized frequency characteristic of $x_n(t)$ as

$$\begin{aligned} Ga_{[x(t)]}(t,f) &= \int_{-\infty}^{\infty} x(\tau) g_w(t-\tau) e^{-j2\pi\tau f} \mathrm{d}\tau \\ &= \int_{-\infty}^{\infty} x(\tau) \left(\frac{1}{\pi\alpha_w}\right)^{1/4} e^{-\frac{(t-\tau)^2}{2\alpha_w}} e^{-j2\pi\tau f} \mathrm{d}\tau \end{aligned} \tag{5}$$

The Gabor transformation of single Gaussian pulse in (1) is

$$\begin{aligned} Ga_{[f(t)]}(t,f) = \sqrt{\frac{2(\alpha\alpha_w)^{1/2}}{\alpha+\alpha_w}} \exp\left[-\left(\frac{\alpha\alpha_w(2\pi f)^2}{2(\alpha+\alpha_w)} + \frac{\alpha\alpha_w t^2}{2(\alpha+\alpha_w)}\right)\right] \\ \bullet \exp\left(-j\frac{2\alpha\pi f t}{\alpha+\alpha_w}\right) \end{aligned} \tag{6}$$

The spectrum is

$$\left|Ga_{[f(t)]}(t,f)\right|^2 = \frac{2\sqrt{\alpha\alpha_w}}{\alpha+\alpha_w} \exp\left[-\left(\frac{\alpha\alpha_w t^2}{\alpha+\alpha_w} + \frac{\alpha\alpha_w(2\pi f)^2}{\alpha+\alpha_w}\right)\right] \tag{7}$$

Gabor transform can optimally reduce the time-frequency window area, and have a good concentration of time and frequency. Therefore, the time and frequency resolutions are compromised, and the best capture of signal characteristics in time-frequency domain is obtained. Another reason that Gabor takes $g_w(t)$ as a Gaussian function is that the Fourier transform of it is still a Gaussian function, which makes the inverse Fourier transform of it is also localized by the window function in frequency domain. Therefore, Gabor transform can achieve the purpose of time-frequency localization: it can

provide all the information of whole signal meanwhile show the intensity changes of signal in any partial time snippet. In short, Gabor transform providing localization details both in time and frequency domain is really meaningful time-frequency analysis.

The TVF principle is to set the time-varying pass region R in (t, f) plane as a mask to trim low energy area to zero while remaining high energy concentration area is expressed as $M(t, f)$, which is defined 1 in R and 0 outside. $M(t, f)$ is calculated according to TFD of original transmitted signal and predetermined threshold r.

$$M(t,f) = \begin{cases} 1 & E/E_{\max} \geq r \\ 0 & E/E_{\max} < r \end{cases} \tag{8}$$

3 Time-Varying Filter Design

The time-varying filter is designed with implicit method summarized into three steps: analysis-mask-synthesis as shown in Fig. 5. Firstly we get the TFD $R(t, f)$ of receiver signal $r(t)$, and use the time-frequency masking function $M(t, f)$ to filter $R(t, f)$ and obtain $R'(t, f)$, then do inverse transform $R'(t, f)$ into $r'(t)$ in time domain, completing the whole process of filtering.

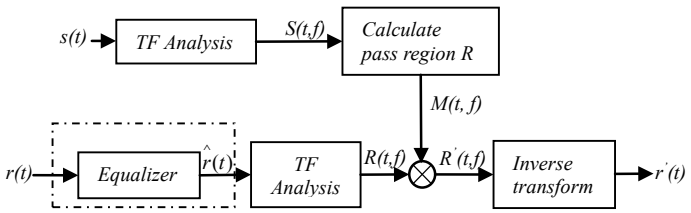


Fig. 5. TVF process at transmitter and receiver

The optimal threshold r_{opt} here is determined by maximum output SINR criterion.

$$\begin{aligned} r_{opt} &= \arg\max\{SNR_{out}\} \\ &= \arg\max_{0<r<1}\left\{10\lg\frac{E\left[|x(n)|^2\right]_{x\in R}}{E\left[\left|\sum i(n)\right|^2\right] + E\left[|w(n)|^2\right]_{i,w\in R}}\right\} \end{aligned} \tag{9}$$

When in fading channel we propose the implementation of joint filter method combining channel equalizer filter with TVF, inhibit ISI/ICI influenced by fading channel, where the TFD is calculated according to equalized signal $\hat{r}(t)$ as shown in the dotted box in 0 then perform the TVF. When the channel situation information (CSI) is known, the frequency domain equalization can be used directly. If CSI is unknown,

it will be estimated by addictive unique word (UW) as pilot frequency. The reception UW sequence divides the transmitted corresponding UW so that we can obtain the response in frequency domain and interpolate to acquire full frequency response. The coefficient of equalizer filter is calculated by the minimum mean square error (MMSE) principle in frequency domain.

4 Simulation and Analysis

In this chapter, we have simulated the linear time-varying filter (LTV) in the resource allocation model shown in Table 1. Concretely, the signal is superimposed with two adjacent frequency interference signals delayed half symbol duration.

In simulation, signals are modulated by BPSK, in which useful signal is f_c = 60 MHz and T_b = 0.26042 μs shaped in Gaussian pulse, while the adjacent interference signals differ by Δf delayed half of T_b. Co-channel interference is shaping in Rectangle pulse.

The parameters of fading channel including the max Doppler frequency shift f_{dmax}, multipath number *NUM*, multipath delay τ, and power loss *pdB* in 0. The channel 1 refers to flat fading channel, while channel 2–4 refer to fast fading channel, frequency selective fading channel, and double selective fading channel.

Table 1. Simulation parameters of fading channel

	f_d/Hz	NUM	τ/s	pdB/dB
Channel 1	20	5	[0 0.13 0.5 1.2 2.1]/fs	[0 −0.967 −1.933 −3.86 −4.84]
Channel 2	4e6	5	[0 0.13 0.5 1.2 2.1]/fs	[0 −0.967 −1.933 −3.86 −4.84]
Channel 3	20	5	[0 10 50 250 450]/fs	[−1 −1.3 −1.5 −1.65 −1.8]
Channel 4	4e6	5	[0 10 50 250 450]/fs	[−1 −1.3 −1.5 −1.65 −1.8]

When in simulation with SIR_{in} = 0 dB, the energy threshold of the time-varying filter is r_{opt} = 0.45 and the mask of LTV filter is designed according to the TFD of {1, 0, 1, 0, 1, 0, 1, 0, 1, 0}. As comparison, the LTI filter is designed by Butterworth bandpass filter with width 4 MHz. The simulation under AWGN channel with SIR_{in} = 0 dB is shown in 0. The BER simulation under fading channel 1–4 is shown in Figs. 6 and 7.

In 0, when in the dense resource allocation system proposed, BER performance of LTV filter is obviously superior to LTI filter. This is due to the pass region of LTV filter remaining most energy of desired signals and filtering most interference energy efficiently, while LTI filter collect signals with more interference. Additionally with Δf declining, BER performances both get worse, but advantage of LTV filter is still clear, because guard band is compressed and ICI gets serious. In 0, the joint filter can suppress interference in fading channel, because MMSE equalizer corrects phase and amplitude distortion suppressing ICI/ISI and improve LTV performance.

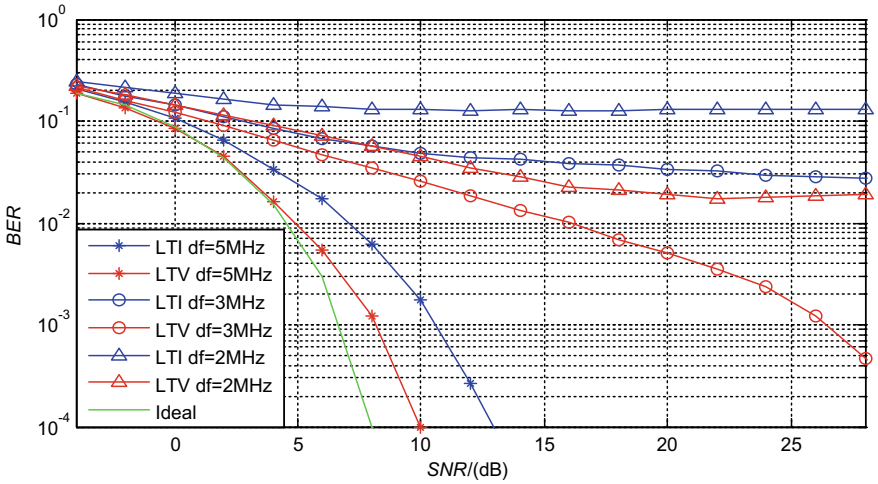


Fig. 6. LTV filter and LTI filter under AWGN channel

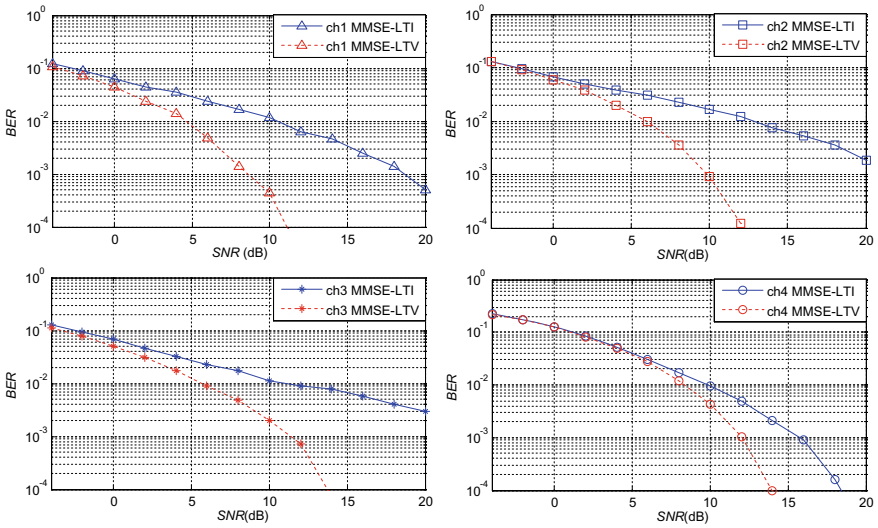


Fig. 7. LTV filter and LTI filter under fading channel

5 Conclusion

In this paper, we analyzed the feasibility of dense resource allocation system model utilizing low energy area of pulse shape signals in (*t*, *f*) plane, in which time-varying filter is used to suppress the interference. At last, we verified the BER performance of time-varying filter under AWGN channel, and the joint filter combined equalizer with time-varying filter in fading channel, which can get better BER performance.

Acknowledgment. This work was supported by the National Nature Science Foundation of China under Grant 61671179.

References

1. Osseiran, A., Boccardi, F., Kusume, K., Maternia, M., Marsch, P., Braun, V., et al.: Scenarios for 5G mobile and wireless communications. IEEE Commun. Mag. **52**(5), 26–35 (2014)
2. Li, Y., Sha, X., Ye, L.: An SC-FDMA inter-cell interference suppression method based on a time-varying filter. In: IEEE Conference on Computer Communications Workshops (2016)
3. Boashash, B.: Time-Frequency Signal Analysis and Processing, p. 2016. Academic Press, Switzerland, Europe (2016)
4. Thomas, S., Scott, B.: Optimal OFDM design for time-frequency dispersive channels. IEEE Trans. Commun. **51**(7), 1111–1122 (2003)
5. Yang-Seok, C., Voltz, P.J., Cassara, F.A.: On channel estimation and detection for multicarrier signals in fast and selective Rayleigh fading channels. IEEE Trans. Commun. **49** (8), 1375–1387 (2001)
6. Ouyang, X., Amin, M.G.: Short-time Fourier transform receiver for nonstationary interference excision in direct sequence spread spectrum communications. IEEE Trans. Signal Process. **49** (4), 851–863 (2001)
7. Cohen, L.: Time-Frequency Analysis, pp. 44–110. Prentice-Hall PTR (1995)
8. Qian, S., Chen, D.: Joint time–frequency analysis. Signal Process. Mag. **16**(2), 52–67 (1999)

Fast Mode Decision of Inter-Prediction for Enhancement Layer in SHVC

Xin Lu(✉), Chang Yu, Xuesong Jin, and Yanfeng Gu

School of Electronics and Information Engineering,
Harbin Institute of Technology, Harbin 150001, China
Xin@hit.edu.cn

Abstract. In the scalable extension of H.265/HEVC, the inter-mode decision process of the enhancement layer (EL) results in greatly increased computing complexity. A fast inter-mode decision algorithm for Scalable High Efficiency Video Coding (SHVC) is proposed to achieve early termination for inter-mode selection based on temporal dependency and motion activity. Statistical results indicate that a Prediction Unit (PU) in the EL is more likely to be coded with the same mode as its temporally co-located PU in the same layer, especially when the movement of the PU is slow. Based on an adaptive double thresholds scheme, the motion complexity can be categorized to three types and the number of inter-mode candidate can be reduced. Simulation results show that the encoding time reduction can be up to 47.29% compared with the original SHM encoder, while the degradation in coding efficiency is acceptable.

Keywords: SHVC · Enhancement layer · Inter-mode · Adaptive thresholds · Temporal correlation

1 Introduction

SHVC standard is developed by the Joint Collaborative Team on Video Coding (JCT-VC) and it is the extension of High Efficiency Video Coding (HEVC). SHVC targets fulfill different video transmission demand while maintaining a high compression efficiency. It inherits advanced technologies that have been used in HEVC. However, these technologies result in greatly increased computational complexity, which make the real-time implementation of SHVC an obstacle.

SHVC standard employs the identical coding structures as HEVC, including Coding Unit (CU), PU, Transform Unit (TU). 8 different inter-modes make the encoding of each CU more effective but complex. When conducting the inter-prediction, all possible partition modes are examined for a CU to select the optimal PU mode based on the Rate Distortion Cost (RDC). Consequently, a more accurate inter-coding in SHVC is achieved at the expense of much higher computational complexity.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 1036–1044, 2019.
https://doi.org/10.1007/978-981-13-6264-4_122

Several methods have been proposed to accelerate the process of inter-mode decision. In [1], Wang et al. suggested an algorithm in which the spatiotemporal characteristics of the encoded frames were used, and a CU depth termination scheme based on the texture complexity was designed. A fast inter-mode decision algorithm was described in [2], in which the distribution of the probability for each inter-mode between the BL and the EL was analyzed.

In this paper, a fast inter-mode decision algorithm is proposed to decrease the overall computational complexity for SHVC inter-encoder. Since the compression process of frames in the EL takes up most of the encoding time, an efficient inter-coding algorithm for the EL is critical in real-time application. Considering the relationship of inter-modes between the current PU and its co-located PU in the previously temporal adjacent frame, the number of possible modes can be reduced first. Then in combination with the motion homogeneity of the current PU, the best mode will be decided. As the thresholds play an crucial role in classifying the motion homogeneity, the adaptive updating algorithms are described in this paper. Experimental results show that the proposed algorithm can significantly reduce the computational complexity of the inter-prediction process with a acceptable compression efficiency loss.

2 Inter-Frame Coding in SHVC

As shown in Fig. 1, SHVC offers 8 PU partitions for inter-coding, namely $2N \times 2N$, $2N \times N$, $N \times 2N$, $N \times N$, $2N \times nU$, $2N \times nD$, $nL \times 2N$, and $nR \times 2N$, where N is the half of the largest CU size. In the inter-prediction process of SHVC, each CU should be encoded by one of the 8 PU partitioning modes. All the 8 modes are evaluated according to the rate distortion function and the mode with minimum RDC is selected to encode the current CU.

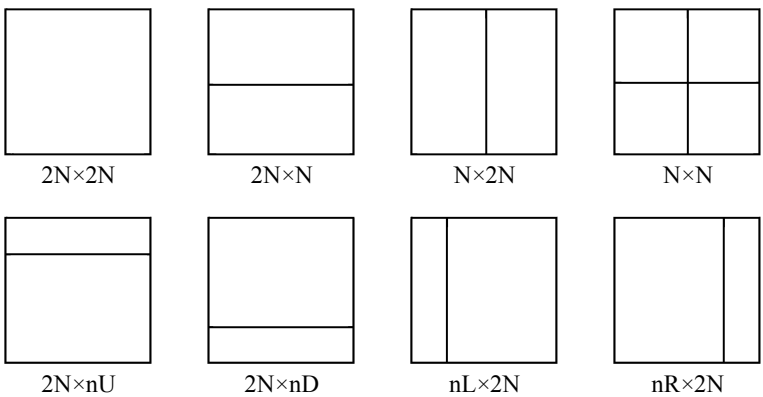


Fig. 1. PU modes in SHVC

3 Proposed Algorithm for Enhancement Layers

3.1 Temporal Correlation of Inter-Mode

Since successive frames tend to have similar content, there is a strong correlation between the current frame and the previously encoded frame. In order to determine the temporal dependency of inter-mode, extensive experiments were performed to measure the probability that the current PU in the EL is coded with the same mode as its co-located PU in the previous frame. In the experiment, the results obtained from three different video sequences with different spatial resolution and varying motion complexities were averaged under the Quantized Parameters (QP) of 32 and 30 for the BL and EL, respectively. The frames in the first two seconds were encoded for each video sequence. Thus, a large number of samples was used to prove the temporal correlation.

Table 1. The relationship of inter-mode between the current PU and its co-located PU in the previous frame

	$2N \times 2N$	$2N \times N$	$N \times 2N$	$N \times N$	$2N \times nU$	$2N \times nD$	$nL \times 2N$	$nR \times 2N$
$2N \times 2N$	0.9518	0.0116	0.0198	0.0050	0.0022	0.0020	0.0043	0.0034
$2N \times N$	0.9340	0.0195	0.0303	0.0017	0.0032	0.0032	0.0044	0.0037
$N \times 2N$	0.9431	0.0146	0.0251	0.0020	0.0041	0.0025	0.0044	0.0042
$N \times N$	0.7375	0.0035	0.0093	0.2446	0.0009	0.0016	0.0009	0.0017
$2N \times nU$	0.9539	0.0143	0.0186	0.0008	0.0041	0.0002	0.0043	0.0034
$2N \times nD$	0.9214	0.0237	0.0398	0.0011	0.0040	0.0037	0.0061	0.0038
$nL \times 2N$	0.9319	0.0211	0.0311	0.0010	0.0038	0.0037	0.0030	0.0046
$nR \times 2N$	0.9410	0.0209	0.0224	0.0016	0.0049	0.0032	0.0025	0.0028

Table 1 illustrates the conditional probability of inter-mode between the co-located PU and the current PU. It can be observed in Table 1 that when coding the current frame in the EL, some unlikely inter-modes can be skipped according to the mode information obtained from the neighboring encoded frame so that the computational complexity can be reduced. For example, when the co-located PU in the previous frame was encoded with $2N \times N$, the probability that the optimal mode of the current PU is selected from $2N \times 2N$, $2N \times N$, $N \times 2N$, $2N \times nU$ and $2N \times nD$ is 98.98%, which means that RDC calculation for the other three inter-modes ($N \times N$, $nL \times 2N$ and $nL \times 2N$) can be avoided while maintaining the high prediction accuracy.

Let $M0$, $M1$, $M2$, $M3$, $M4$, $M5$, $M6$, and $M7$ denote the inter-mode of $2N \times 2N$, $2N \times N$, $N \times 2N$, $N \times N$, $2N \times nU$, $2N \times nD$, $nL \times 2N$, and $nR \times 2N$, respectively. Based on the observation, the initial inter-mode candidate list PR for the current PU in the EL can be determined as

$$PR = \begin{cases} \{M0, M1, M2, M4, M5\} & \textbf{if } M = Ml \\ \{M0, M1, M2, M6, M7\} & \textbf{if } M = M2 \\ \{M0, M1, M2, M3\} & \textbf{if } M = M3 \\ \{M0, M1, M2, M4\} & \textbf{if } M = M4 \\ \{M0, M1, M2, M5\} & \textbf{if } M = M5 \\ \{M0, M1, M2, M6\} & \textbf{if } M = M6 \\ \{M0, M1, M2, M7\} & \textbf{if } M = M7 \\ \{PR0\} & \textbf{if } M = M0 \end{cases} \quad (1)$$

where M represents the mode of the corresponding PU in the previous frame. Due to the fact that more than 90% of PUs in the EL are encoded with inter-mode $2N \times 2N$, when M is $M0$, further improvement should be made to construct the candidate list $PR0$, thus avoiding to examine all the three possible modes ($M0$, $M1$ and $M2$).

3.2 Motion Homogeneity of the PU

For PUs of which the temporal co-located PU was encoded with M0, motion homogeneity was evaluated to further accelerate the inter-mode prediction process. We calculate the motion vector MV, which is defined by Eq. (2) to measure the motion homogeneity of PU.

$$MV = [(MV_H)^2 + (MV_V)^2]^{\frac{1}{2}} \quad (2)$$

where MV_H and MV_V denote the horizontal and vertical motion vectors of the 4×4 block in the current PU while MV_V represents the vertical motion vector.

Generally, PUs containing complex motion result in large MV values, and this means that the video content in the current PU differs greatly from the content in the co-located PU, thus more candidate modes should be included in the list $PR0$. In contrast, PUs with small MV value can be considered as located in an area with regular motion and the optimal inter-mode can be determined as $M0$, namely the same mode of the temporal co-located PU.

3.3 Adaptive Double Thresholds Method

An adaptive double thresholds method is introduced to categorize the motion homogeneity of a PU in the EL based on the MV value. If the MV value of a PU is smaller than the lower threshold T_L, the PU is considered in an area of high motion homogeneity; when the MV value is larger than the upper threshold T_U, it is considered to be located in a region of complex motion and all three candidate modes should be tested to maintain a high prediction accuracy. Otherwise, the PU is considered to be located in a region with regular motion activity.

The coding depth level of the current PU is employed to further reduce the encoding time. For the current PU with MV value larger than T_U, if its coding depth level is greater than 1, which means that rich details appears in the PU,

evaluations for all the three candidate modes ($M0, M1, M2$) are required; if the coding depth level is lower than 2, the texture complexity of the current PU is judged as low and $M0$ can be directly selected as the best mode. When the MV value is between T_L and T_U, if the coding depth level of the current PU is 3, $PR0$ should contain $M0$, $M1$ and $M2$; Otherwise, only the RD calculation for mode $M0$ is needed so that examinations for other modes can be avoided. This process can be summarized as

$$PR0 = \begin{cases} PR_{b1} & \textbf{if } MV > T_U \\ PR_{b2} & \textbf{if } T_L \leq MV \leq T_U \\ M0 & \text{otherwise} \end{cases} \tag{3}$$

$$PR_{b1} = \begin{cases} \{M0, M1, M2\} & \textbf{if } d > 1 \\ \{M0\} & \text{other} \end{cases}, PR_{b2} = \begin{cases} \{M0, M1, M2\} & \textbf{if } d = 3 \\ \{M0\} & \text{other} \end{cases} \tag{4}$$

where d represents the coding depth level of the current PU.

3.4 Selection of Thresholds

If the best mode of corresponding PU of the current PU in the previous neighboring frame is $M0$, by combining MV value with coding depth level, inter-mode candidates for the current PU can be decided by the motion information and the coding depth level. The motion activity can be categorized by comparing the MV value with the thresholds. A selection of appropriate MV thresholds can achieve a tradeoff between prediction accuracy and speed of inter-mode decision, which plays an important role in the fast inter-mode selection algorithm.

In our algorithm, the frames in the EL are categorized into the frames for updating thresholds frames (F_U) and implementing algorithms frames (F_I). The determination and update of double thresholds are implemented in F_U while the determination of the best inter-mode is accomplished in F_I. In F_U frames, the MV value is compared with the double thresholds. If the number of CTU in a frame in the EL is less than 25×15, the video is regarded as a high-resolution video, and the initial values for the upper and lower thresholds are set to 30 and 10, respectively, in the first F_U frame. Otherwise, the current video is categorized as a low-resolution video, and T_U and T_L are set to 0 and 10 respectively.

After coding a F_U frame, and error ratios for all PUs that the co-located PUs in the previous frame are coded with $M0$ are used to evaluate the two thresholds, as calculated by

$$P_{E_U} = E_U/(E_U + C_U), \quad P_{E_L} = E_L/(E_L + C_L) \tag{5}$$

where P_{E_U} and P_{E_L} are error ratios for the upper and lower threshold respectively; E_U is the number of PUs coded with inter-mode other than $M0$, of which the MV value is greater than T_U and the coding depth level is 2 or 3; C_U is the number of PUs coded with $M0$, while its MV value is greater than T_U and the coding depth level is greater than 1. Similarly, E_L represents the number of

PUs not coded by mode $M0$, of which the MV value is lower than T_L and the coding depth level is equal to 3. C_L denotes the number of PUs coded with $M0$ and its MV value is lower than T_U under the coding depth level 3. The initial values are set to 0. After encoding a F_U frame, the values of E_U, E_L, C_U, and C_L are calculated.

In the subsequent F_U frames, if P_{E_U} for the current upper threshold T_U is greater than the acceptable maximum error ratio (P_{max}), the new T_U value is set to T_U -2. On the contrary, when T_U is less than or equal to P_{max}, the updated upper threshold is set to $T_L + 2$.

The procedure of the proposed adaptive upper threshold scheme for the EL are described in Algorithm 1. The lower threshold T_L is updated in a similar way as T_U.

4 Overall Proposed Algorithm

Due to the fact that processing the EL consumes the majority of the overall encoding time in SHVC, our fast algorithm is only used for the EL, which can achieve coding time reduction with a high coding efficiency. If the current frame is a F_I in the EL, the optimal inter-mode of PUs can be fast and accurately

Algorithm 1. Determination of the upper threshold T_U

Input: F_U andC_U and E_U and MV and $uiDepth$ and $uiPreMode$ and $uiCurMode$ and $NumOfLayer$ and $NumOfPU$ and T_U
Output: T_U

```
if NumOfLayer > 0 then
    if F_U == true then
        for j = 0; j < NumOfPU; j++ do
            if MV > T_U and uiDepth > 1 then
                if uiPreMode = M0 and uiCurMode = M0 then
                    C_U++;
                end
                if uiPreMode = M0 and uiCurMode ≠ M0 then
                    E_U++;
                end
            end
        end
        P_{E_U} = E_U ÷ (E_U + C_U);
        if P_{E_U} <= 0.0625 then
            T_U = T_U + 2;
        else
            T_U = T_U − 2;
        end
    end
    return T_U;
end
```

predicted according to the temporal correlation and motion homogeneity. The overall proposed fast mode decision algorithm for inter-prediction in the EL is described in Algorithm 2 in detail.

5 Experimental Evaluation

To evaluate the performance of the proposed fast algorithm, 16 standard test video sequences with different spatial resolution and motion degree are processed in the SHM 12.0 reference software. A two-layer spatial scalability structure is adopted with Low Delay (LD) configuration. The GOP (Group of Picture) is set to 4, and the size of coding tree block is 64×64. The quantized parameters of base layer are 22, 27, 32 and 37, while the corresponding QP factor of the EL is 20, 25, 30, and 35, respectively. Three parameters, $BDPSNR(dB)$, $BDBR(\%)$, as defined in [3], and Time Reduction (TR,%), are employed to validate the effectiveness of the proposed scheme in this paper.

TR is calculated as $TR = ((T_P - T_{SHM})/T_P) \times 100\%$, where T_{SHM} and T_P respectively represent the overall execution time consumed by the SHM 12.0 scheme and the proposed algorithm for encoding both layers.

Table 2 shows the comparison of simulation results. As can be seen in Table 2, compared with the original SHM encoder, the proposed algorithm achieves an

Table 2. Comparision of proposed algorithm and SHM12.0

Class	Sequence	Proposed Algorithm		
		TR	BDBR	BDPSNR
A	PeopleOnStreet	40.53	2.7	−0.12
	Traffic	47.29	2.94	−0.09
B	BasketballDrive	28.5	1.62	−0.03
	BQTerrace	34.08	1.6	−0.03
	Cactus	39.23	2.76	−0.06
	Kimono1	29.44	2.1	−0.07
C	ParkScene	37.28	2.09	−0.06
	BQMall	28.45	2.13	−0.08
	PartyScene	31.89	1.47	−0.07
	RaceHorses	25.49	1.29	−0.06
D	BasketballPass	32.56	1.27	−0.06
	BlowingBubbles	35.45	2.26	−0.09
	BQSquare	37.3	2.17	−0.09
	RaceHorses	30.62	1.51	−0.08
E	FourPeople	33.89	2.6	−0.07
	Johnny	33.67	3.67	−0.07
	Average	**34.10**	**2.13**	**−0.07**

Algorithm 2. Overall Proposed algorithm

Input: MV and T_U and T_L and $NumOfLayer$ and $uiDepth$ and $uiPreMode$
Output: $PR0$;

```
while EndOfCodingCurrentFrame == Flase do
    if NumOfLayer > 0 then
        if F_I == True then
            if uiPreMode == M0 then
                if MV > T_U then
                    if uiDepth > 1 then
                        PR0 = {M0, M1, M2}
                    else
                        PR0 = {M0}
                    end
                end
                if MV < T_L then
                    PR0 = {M0}
                end
                if MV ≥ T_L and MV ≤ T_U then
                    if uiDepth == 3 then
                        PR0 = {M0, M1, M2}
                    else
                        PR0 = {M0}
                    end
                end
            end
            if uiPreMode == M1 then
                PR0 = {M0, M1, M2, M4, M5}
            end
            if uiPreMode == M2 then
                PR0 = {M0, M1, M2, M6, M7}
            end
            if uiPreMode == M3 then
                PR0 = {M0, M1, M2, M3}
            end
            if uiPreMode == M4 then
                PR0 = {M0, M1, M2, M4}
            end
            if uiPreMode == M5 then
                PR0 = {M0, M1, M2, M5}
            end
            if uiPreMode == M6 then
                PR0 = {M0, M1, M2, M6}
            end
            if uiPreMode == M7 then
                PR0 = {M0, M1, M2, M7}
            end
        end
        if F_U == True then
            update T_U and T_L;
        end
    end
end
```

average time reduction of 34.10% regardless of the resolution of video sequence, and the maximum reduction is up to 47.29%. Meanwhile, the proposed fast inter-mode decision algorithm shows a similar rate distortion performance compared with SHM implementation, where the average PSNR decrease is 0.07 dB and the average rate increment is 2.13%.

6 Conclusion

In this paper, a fast inter-mode decision algorithm for inter-coding of the EL in SHVC was proposed based on the optimal coding mode of the co-located PU in the previous neighboring frame and motion degree of the current PU. An adaptive double threshold method was introduced to categorize the motion homogeneity. In comparison with the original SHM encoder, experimental results showed that our method can achieve significant encoding time reduction, which is up to 47%, while the loss of coding efficiency is acceptable.

References

1. Wang, X., Xue, Y.: Fast HEVC inter prediction algorithm based on spatio-temporal block information. In: IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), vol. 2017, pp. 1–5. Cagliari (2017)
2. Kim, S.T., Konda, K.R., Park, C.S., Cho, C.S., Ko, S.J.: Fast mode decision algorithm for inter-layer coding in scalable video coding. IEEE Trans. Consum. Electron. **55**(3), 1572–1580 (2009)
3. Bjontegaard, G.: Calculation of average PSNR differences between RD-curves. ITU SG16 Doc. VCEG-M33 (2001)

An Adaptive Spectrum Sharing Scheme Based on Game Theory for Millimeter Wave RoF Systems

Na Chen(✉), Xiaoguang Zhang, and Songlin Sun

Beijing University of Posts and Telecommunications, Beijing, China
chennabupt@163.com

Abstract. The millimeter wave (mmWave) radio over fiber (RoF) system is a promising solution for future communication systems. This paper focuses on the spectrum sharing problem in a mmWave RoF system. We consider a heterogeneous network (HetNet) model and propose an adaptive scheme. The scheme identifies the users as licenced users that primarily owns the specific spectrum and unlicenced users that have to purchase the spectrum. Simulation results finally show that the proposed spectrum sharing scheme can improve the spectrum efficiency.

Keywords: RoF · mmWave · Game Theory

1 Introduction

For the upcoming modern communication system such as 5G, seamless access as well as high data rate up to 10–20 Gbps are expected. To follow the trend of the application of millimeter wave (mmWave) [1], radio over fiber (RoF) technologies can provide high bitrate mmWave signals with optical multilevel modulation techniques [2–4]. However, for RoF systems, there is still an tough issue on the spectrum shortage of the signals [5,6].

This paper focuses on the spectrum sharing in a mmWave RoF heterogeneous network (HetNet). The large bandwidth makes mmWave as a promising technology to provide high data rates for small cells. And the physical character of the short wavelength makes mmWave highly directional with narrow beams and sensitive to blockage [7]. This results in lower interference and an effective usage of the spectrum. However, within a relatively small area, the conflict still exists which lower the quality of service (QoS). So the spectrum sharing should be more feasible for this scenario.

Here we summarize several outstanding related works as follows. Reference [7] proposed an optimized scheme by setting the interference threshold for the secondary licenced network. Reference [8] proposed a full-duplex quasi-gapless carrier aggregation scheme for RoF HetNets. Reference [9] focused on the seamless integration of the RoF system and evaluated 5 Gbps signal transmission. Authors of [10] researched on the optical frequency combs as a promising solution

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 1045–1052, 2019.
https://doi.org/10.1007/978-981-13-6264-4_123

for the efficient mmWave generation method. References [11,12] proposed multi-scale approaches to spectrum sensing and information exchange in mmWave cognitive cellular networks with small cells.

We summarize the main contributions of the work as follows. Formulate an improved Stackelberg game spectrum sharing scheme to increase the spectrum efficiency. At the beginning, the eNB of the licensed network price the shared spectrum. If the LPNs would like to transmit on some extra band, they need to purchase it by paying the corresponding price. However, there exists the original spectrum hole ratio of the eNB coverage. By identifying the spectrum hole, the LPN can decide whether to purchase extra bandwidth and the number of licensed nodes to offload. By deploying the dynamic scheme, balance is achieved with the most suitable parameters generated by the algorithm.

The rest of the paper is organized as follows. We introduce the system model and presents the basic idea of the system in Sect. 2. In Sect. 3, we propose and analyze the adaptive scheme. Section 4 presents the simulation setup as well as the numerical results, then discusses the system performance. Finally, in Sect. 5, the paper is concluded.

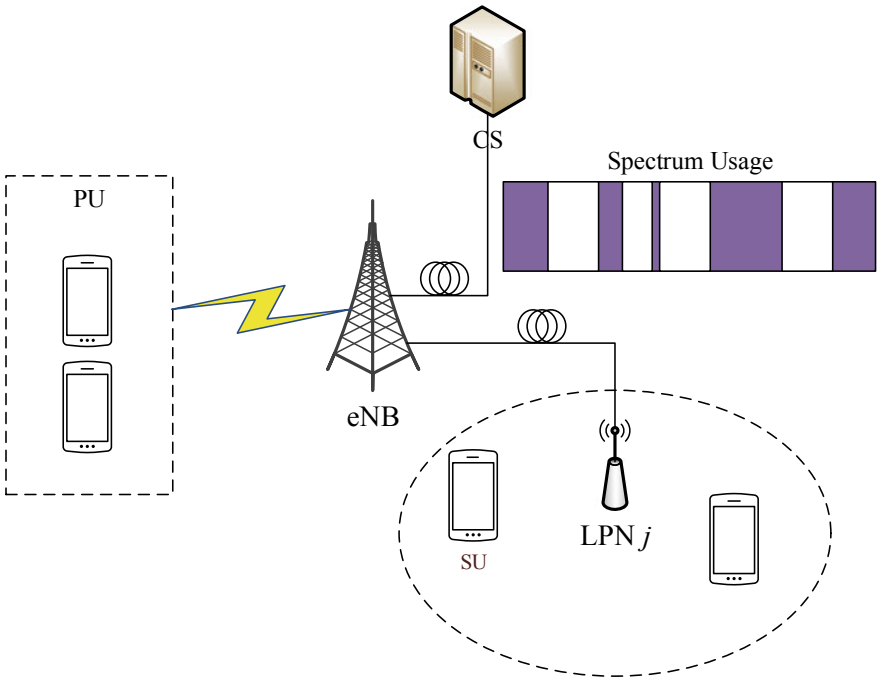


Fig. 1. The system structure.

2 System Model

We consider the mmWave RoF system as is shown in Fig. 1. The central station (CS) generates optical signals and transmit to the eNB who serves the primary users (PUs). Then within the macro cells there are multiple small cells including secondary users (SUs) served by low power nodes (LPNs).

Suppose there are J co-channel LPNs in a single macro cell. The set of the eNB and LPNs is presented by $j \in \mathcal{B} = \{0, 1, \ldots, J\}$, where $j = 0$ represents the eNB and $j \geq 1$ stand for the LPN indexed j. And the set of I external interference nodes are denoted by $i \in \mathcal{B}_I = \{1, 2, \ldots, I\}$. N_0 and N_j represents the number of PUs and SUs in the j-th LPN respectively.

Assume the nodes in the network transmit and receive signals in the mmWave band. The channel condition is supposed to be block-fading and the zero mean additive noises are independent circularly symmetric complex Gaussian (CSCG) with variance σ^2.

3 Proposed Scheme

3.1 Compressed Sensing

For the modern communication system, cognitive radio (CR) is capable to make better use of the limited wireless communication resources by detecting the variation of surrounding radio environment and responding to any system change. Ideally, CR can sense broad wide frequency band precisely and response on time. However, it is impractical to sense the wideband spectrum in the complex realistic channel, the spectrum sensing operation causes delay, and sensing the spectrum holes accurately is of high difficulty in practice. To help with solving the problem, the compressed sensing (CS) theory is deployed in CR system to alleviate the system pressure, which is an effective frame especially for wideband spectrum sensing.

CS aims to recover high dimensional sparse signals with considerably fewer dimensions, which is expressed as

$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{w} = \mathbf{\Phi \Psi \theta} + \mathbf{w} = \mathbf{\Theta \theta} + \mathbf{w}, \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^N$ is a sparse vector estimated from an observed channel spectrum state vector $\mathbf{y} \in \mathbb{C}^M$, and $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$ is a measurement matrix. $\mathbf{w} \in \mathbb{C}^N$ represents the noise. However, the signal to be recovered is usually not sparse in practical applications, but can be represented in some basis $\{\mathbf{\Psi}\}_{i=1}^N$ with the corresponding sparse coefficients θ_i. $\boldsymbol{\theta} \in \mathbb{C}^L$ is an L dimension sparse vector of coefficients based on a basis matrix $\mathbf{\Psi} \in \mathbb{C}^{N \times L}$, and $\mathbf{\Theta} = \mathbf{\Phi \Psi}$ is an $M \times L$ matrix ($M \ll L$).

We utilize l_0 norm minimization for the reconstruction process to obtain the sparsest sequence among all the available solutions.

$$\widehat{\boldsymbol{\theta}} = min \|\boldsymbol{\theta}\|_0 \ s.t. \ \|\mathbf{\Theta \theta} - \mathbf{y}\|_2 \leq \Xi, \tag{2}$$

where $\widehat{\boldsymbol{\theta}}$ denotes the estimated vector for $\boldsymbol{\theta}$ and $\|\mathbf{w}\| \leq \Xi$ is the noise tolerance.

Obviously, only when matrix $\boldsymbol{\Theta}$ cannot map two different S-sparse signals to the same set of samples, can the reconstruction of a compressible signal $\mathbf{x}$ be reliable, implying the matrix $\boldsymbol{\Theta}$ should satisfy the restricted isometry property (RIP).

3.2 Stackelberg Game Formulation

Let r^0 be the minimum data rate demanded for a random power node, while r is the true minimum user data rate defined for the licensed network. As for a node among N attached nodes, we have $U = N \ln\left(\frac{r}{r^0}\right)$ as the general utility function.

The LPNs aim to obtain maximum data rates with the least cost, their utility function is defined as

$$U_j = CN_j \ln\left(\frac{r_j}{r^0}\right) - (1-\gamma)\, AF_j, \tag{3}$$

where C is a positive constant, and A is the unit price of the frequency band. To optimize the LPN j, we have

$$\max_{F_j}\ U_j \quad st. 0 \le F_j \le F. \tag{4}$$

The utility function for the licensed network is

$$U_M = \sum_{j\ge 1} (1-\gamma)\, AF_j + C\left(N_M \ln\left(\frac{r_M}{r^0}\right) + \sum_{j\ge 1} N_j^{offload} \ln\left(\frac{r_j^{offload}}{r^0}\right)\right). \tag{5}$$

For the licensed network, formulate the optimization problem as

$$\max_{\gamma,\boldsymbol{\beta}}\ U_M st.\ 0 \le \gamma \le 1 \quad \beta_j \ge 1,\ j = 1, 2, ..., J, \tag{6}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3 ... \beta_F)$.

The eNB of the licensed network randomly initializes its parameters γ and $\boldsymbol{\beta}$. LPNs decide whether to purchase the bandwidth after solving the optimization problem. For new γ and $\boldsymbol{\beta}$, LPNs purchase some bandwidth. The eNB performs a backward induction to achieve the balance.

Here we have U_j as a function of F_j, with second derivative formulated as

$$\frac{\partial^2 U_j}{\partial {F_j}^2} = \frac{-N_j C}{{F_j}^2} < 0. \tag{7}$$

Then solve the maximal F_j with the given γ

$$\frac{\partial U_j\,(F_j)}{\partial F_j} = \frac{N_j C}{F_j} - (1-\gamma)\, A, \tag{8}$$

The solution F_j^* is obtained as

$$F_j^* = \begin{cases} F, & 1 - \frac{N_j C}{AJ} \le \gamma \le 1 \\ \frac{N_j C}{A(1-\gamma)}, & 0 \le \gamma < 1 - \frac{N_j C}{AJ} \end{cases} \tag{9}$$

Generally, the LPN cannot occupy all the spectrum usage of the licensed network, so $\gamma < 1 - \frac{N_j C}{AJ}$ is feasible in most situation. It is helpful to assume that $F_j^*(\gamma)$ equals to $\frac{N_j C}{A(1-\gamma)}$ to further simplify the problem.

We transform U_M into the form $F_j^*(\gamma)$ for the backward induction.

$$U_M(\gamma, \boldsymbol{\beta}) = C\sum_{j\geq 1} N_j + Cf_0(\boldsymbol{\beta}) \ln\left((1-\gamma) Q_0(\boldsymbol{\beta})\right) + C\sum_{j\geq 1} f_j(\beta_j) \ln\left(\frac{\gamma Q_j(\beta_j)}{1-\gamma}\right), \tag{10}$$

where

$$Q_0(\boldsymbol{\beta}) = \frac{F}{f_0(\boldsymbol{\beta}) r^0} \log\left(1 + SINR_{0,min}^{(0)}\right) \tag{11}$$

$$Q_j(\beta_j) = \frac{N_j C}{A f_j(\beta_j) r^0} \log\left(1 + \frac{SINR_{j,min}^{(1)}}{\beta_j}\right) \tag{12}$$

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3 ... \beta_F). \tag{13}$$

To solve the above-mentioned problem, $\boldsymbol{\beta}$ needs to be dependent. We consider the following two cases: (i) $\boldsymbol{\beta}$ is predetermined; (ii) for each LPN, change $\boldsymbol{\beta}$ until achieving the maximum access data rate. In [13], we gave the detailed solution for $\boldsymbol{\beta}$.

3.3 Performance Analysis

According to previous assumptions, the signal to interference plus noise ratio (SINR) of the user k served by the node j in the non-spectrum hole period is denoted by

$$SINR_{j,k}^{(0)} = \frac{P_j h_{j,k}}{\sum_{s\in\mathcal{B}/j} P_s h_{s,k} + \sum_{i\in\mathcal{B}_I} P_i' h_{i,k}' + \sigma^2}. \tag{14}$$

Meanwhile, the SINR during the spectrum hole period can be expressed as

$$SINR_{j,k}^{(1)} = \frac{P_j h_{j,k}}{\sum_{s\in\mathcal{B}/\{0,j\}} P_s h_{s,k} + \sum_{i\in\mathcal{B}_I} P_i' h_{i,k}' + \sigma^2}. \tag{15}$$

In the Eq. (13), (14), P_j and $P_i^{'}$ represent the transmit power of LPN j and external power node i, respectively. The path gain between the user k and LPN j is expressed by $h_{j,k}$, and $h_{i,k}^{'}$ indicates the path gain between the user k and external power node i. We also assume that the LPNs are sparsely located in the licensed network and the interference between them is ignored for simplicity.

For LPN j, the minimum user data rate before offloading licensed users is

$$r_j = \frac{F_j(1-\gamma)}{N_j} \log(1 + SINR_{j,min}^{(0)}), \tag{16}$$

where γ and F_j express the spectrum hole ratio and the relay j's bandwidth, respectively, and $SINR_{j,min}^{(0)}$ denotes the minimum SINR among N_j users served

by LPN j in the non-spectrum-hole period. Similarly, the minimum user data rate of LPN j after offloading the licensed users can be expressed as

$$r_j^{offload} = \frac{F_j \gamma}{N_j^{offload}} \log(1 + \frac{SINR_{j,min}^{(1)}}{\beta_j}), \tag{17}$$

where β_j is the offloading factor, $SINR_{j,min}^{(1)}$ is the minimum SINR obtained among N_j in spectrum holes. Here, $N_j^{offload} = f_j(\beta_j)$ is an increasing function with the variable β_j, and in practice, we obtain the $f_j(\cdot)$ by analyzing long term statistic with particular distributions of users and LPNs. Hence, the minimum user data rate of the licensed network is denoted by

$$r_M = \frac{F(1-\gamma)}{N_M} \log(1 + SINR_{0,min}^{(0)}), \tag{18}$$

where F represents the total bandwidth of the eNB, $N_M = f_0(\beta_1, \beta_2, \ldots, \beta_J)$ is a nonincreasing function with offloading factors as its variable, and $SINR_{0,min}^{(0)}$ denotes the minimum SINR of the N_M users. Then the relationship between N_M and $N_j^{offload}$ satisfies

$$N_M + \sum_{j=0}^{J} N_j^{offload} = N_0. \tag{19}$$

4 Similation Results

We set up the simulation and present the corresponding numerical results in this section, then evaluate the performance of several schemes including the one proposed. Table 1 lists the main parameters for the simulation.

As is shown in Fig. 2, we compare the coverage probability of the different groups of users. As the SINR threshold increases, the coverage probability

Table 1. Main parameters

Parameters	Values
Simulation scenario	ITU-UMa/UMi
Macro cell radius	500 m
eNB TX power	46 dBm
Small cell radius	50 m
LPN TX power	30 dBm
System bandwidth	100 Mhz
Distribution of users	Uniform
Terminal noise density	−174 dBm/Hz

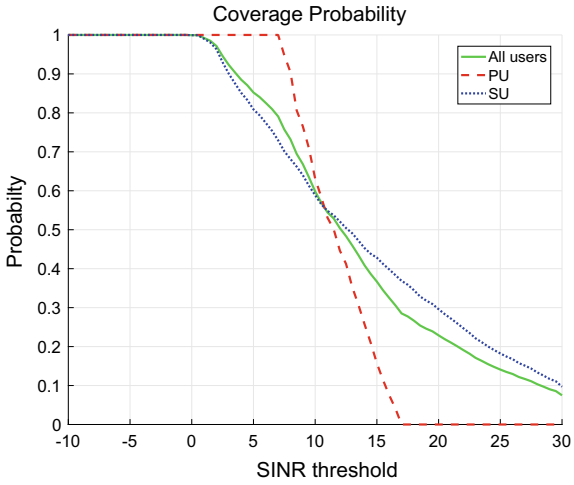
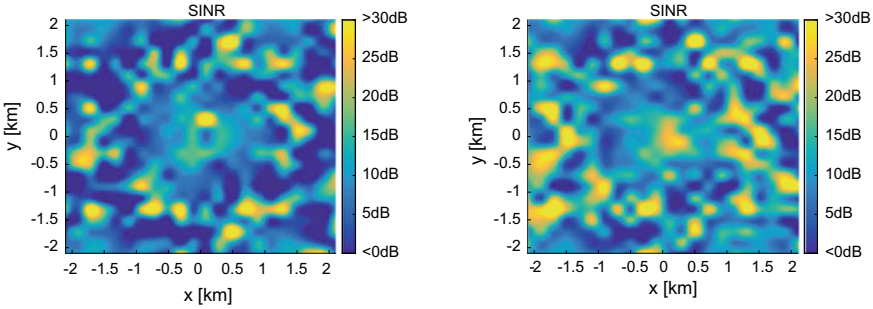


Fig. 2. The coverage probability v.s. the SINR threshold.



(a) SINR distribution without deploying the proposed scheme

(b) SINR distribution when deploying the proposed scheme

Fig. 3. SINR distribution before and after the scheme is deployed.

decreases. We can observe that the coverage probability is higher for PUs than that of SUs at relatively lower SINR threshold, but the ratio becomes lower at higher SINR threshold. One of the possible reasons is that PU is closer to the eNB and may be interfered by LPNs when the SINR threshold becomes higher.

The SINR distribution before and after the schemes is shown in Fig. 3. As we can see, the SINR increases globally, which shows that the proposed scheme is helpful for the system performance on the overall SINR.

5 Conclusions

In this paper, we focused on the spectrum sharing problem and proposed an adaptive spectrum sharing scheme to improve the SINR with spectrum efficiency maintained for the HetNet based on RoF system. First, the paper indicated

the problem of spectrum usage in mmWave RoF systems. Then, we set up the framework of compressive sensing based Stackelberg game algorithm to serve for the BUs in the network. Then, we analyze the performance of the scheme. Finally, we simulated the scheme and it is capable to serve most of the users and improves the system SINR.

Acknowledgment. Project 61471066 supported by NSFC.

References

1. Wang, Y., Pei, L., Li, J., Li, Y.: Millimeter-wave signal generation with tunable frequency multiplication factor by employing UFBG-based acousto-optic tunable filter. IEEE Photon. J. **9**(1) (2017)
2. Beas J., Castanon, G., Aldaya, I., Aragon-Zavala, A., Campuzano, G.: Millimeter-wave frequency radio over fiber systems: a survey. IEEE Commun. Surv. Tutor. **15**(4), 1593–1619 (2013)
3. Li, X., Kong, M., Yu, J., Chang, G.-K.: Optimization of modulation formats for high performance millimetre wave RoF transmission system. In: 2017 European Conference on Optical Communication (ECOC), pp. 1–3 (2017)
4. Kanno, A.: High bitrate mm-wave links using RoF technologies and its non-telecom application. In: 2017 Optical Fiber Communications Conference and Exhibition (OFC), pp. 1–3 (2017)
5. Caballero, A., Borkowski, R., Miguel, I., Durn, R.J., Aguado, J.C., Fernndez, N., et. al: Cognitive, heterogeneous and reconfigurable optical networks: the CHRON project. J. Lightwave Technol. **32**(13), 2308–2323 (2014)
6. Thomas, V. A., El-Hajjar, M., Hanzo, J.: Performance improvement and cost reduction techniques for radio over fiber communications. IEEE Commun. Surv. Tutor. **17**(2), 627–670 (2015)
7. Gupta, A.K., Alkhateeb, A., Andrews, J.G., Heath, R.W.: Gains of restricted secondary licensing in millimeter wave cellular systems. IEEE J. Select. Areas Commun. **34**(11), 2935–2950 (2016)
8. Zhang, J., Xu, M., Wang, J., Lu, F., Cheng, L., Cho, H., et. al: Full-duplex quasi-gapless carrier aggregation using FBMC in centralized radio-over-fiber heterogeneous networks. J. Lightwave Technol. **35**(4), 989–996 (2017)
9. Bekkali, A., Nishimura, K.: Seamless convergence of radio-over-fiber and millimeter-wave links for highly resilient access networks. In: 2016 IEEE Wireless Communications and Networking Conference, pp. 1–6 (2016)
10. Browning, C., Elwan, H.H., Martin, E.P., Duill, S.O., Poette, J., Sheridan, P., et. al: Gain-switched optical frequency combs for future mobile radio-over-fiber millimeter-wave systems. J. Lightwave Technol. (2018)
11. Michelusi, N., Nokleby, M., Mitra, U., Calderbank, R.: Multi-scale spectrum sensing in millimeter wave cognitive networks. In: 2017 51st Asilomar Conference on Signals, Systems, and Computers, pp. 1640–1644 (2017)
12. Michelusi, N., Nokleby, M., Mitra, U., Calderbank, R.: Multi-scale spectrum sensing in small-cell mm-wave cognitive wireless networks. In: 2017 IEEE International Conference on Communications (ICC), pp. 1–6 (2017)
13. Sun, S., Chen, N., Ran, T., Xiao, J., Tian, T.: A stackelberg game spectrum sharing scheme in cognitive radio-based heterogeneous wireless sensor networks. Signal Process. **126**(2016), 18–26 (2016)

Unsupervised Hyperspectral Band Selection Method Based on Low-Rank Representation

Chunyan Yu[1], Kun Cen[1(✉)], Chein-I Chang[1,2], and Fang Li[1]

[1] Dalian Maritime University, Dalian 116026, China
ckun@dlmu.edu.cn
[2] Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Baltimore, MD, USA

Abstract. In order to reduce the spectral redundancy of hyperspectral remote sensing images and reduce the computational complexity of subsequent processing, an unsupervised hyperspectral image band selection algorithm based on low-rank representation (LRBS) was proposed in this paper. First, a low-rank representation of the hyperspectral image is proposed and a low-rank coefficient matrix is obtained. Then, each column of the low-rank coefficient is used as a vertex of the graph to perform spectral clustering. Lastly, we use the fixed initial k-means cluster centers for clustering to get the salient band of each cluster. The experimental simulation results show that the bands selected by LRBS algorithm can improve the classification accuracy and have better performance than other methods.

Keywords: Band selection (BS) · Low-rank representation (LRR) · Hyperspectral image classification

1 Introduction

HYPERSPECTRAL imaging (HSI) is one of the most powerful technologies to remotely detect and recognize the material properties of the object in the interest scene for the broadband wavelength and high spectral resolution [1–3] However, a large number of spectral bands pose a great challenge for information extraction. For example, in hyperspectral image classification, the classification accuracy decreases with the increase of the number of bands relative to a small number of tag samples. Therefore, it is very necessary to reduce the amount of data and save resources. Two methods are selected for dimension reduction: feature extraction and band selection. Feature extraction uses mapping methods to convert raw data to fewer new features including principal component analysis (PCA) and independent component analysis (ICA) [4]. However, feature extraction will change the original data to a certain degree. Different from feature extraction, bands selection selects proper band subsets from the original data set [5]. It not only preserves the original features and physical meaning of remote sensing data but also reduces the size of the data to achieve the goal of dimensionality reduction.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 1053–1061, 2019.
https://doi.org/10.1007/978-981-13-6264-4_124

Recently, some hyperspectral BS methods have been proposed for hyperspectral classification and target detection. Weiwei sunpresented a novel symmetric sparse representation (SSR) method to solve the band selection problem in hyperspectral imagery classification [6]. H Su et al. proposed a particle swarm optimization (PSO)-based system to select bands and determine the optimal number of bands to be selected simultaneously [7]. Low-rank representation (LRR) is a new tool for robustly and efficiently processing high-dimensional data. Xiaotao Wang proposed a predimension-reduction algorithm that couples weighted low-rank representation (WLRR) with a skinny intrinsic mode functions (IMFs) dictionary for hyperspectral image (HSI) classification. Alex Sumarsono et al. introduced the LRR model into the hyperspectral image and used it to estimate the number of subspaces [8–11]. However, these low-rank-based methods do not make full use of the resulting low-rank information.

In this paper, we developed an unsupervised BS approach, to be called low-rank band selection (LRBS) for hyperspectral image processing where no *a priori* knowledge available to be used. The organization of this paper is organized as follows. In Sect. 2, the proposed method is specifically described. Section 3 presents the extensive experiments and conclusions are drawn in Sect. 4.

2 LRBS Method

2.1 The LRR Model

A hyperspectral image is denoted as $X = (x_1, x_2, \cdots, x_L)^T \in R^{L\times n}$, where L is the number of bands, n is the total number of pixels, and $\boldsymbol{x_i} = (x_{i,1}, x_{i,2}, \cdots, x_{i,n})^T$ is the i^{th} spectral band. The HSI data includes the low-rank target parts and the noise part. Therefore, the following model is defined as low-rank representation of HSI:

$$\min \Gamma(\mathbf{Z}) + \lambda \|\mathbf{E}\|_{1,2}, \quad st. \quad X = DZ + E, \tag{1}$$

where $\Gamma(\bullet)$ is the coefficient of the low-rank matrix, D is the dictionary of HSI, E denotes the noise, λ is the balance parameter of the two parts of the equation. You can get λ by

$$\lambda = \frac{1}{\sqrt{\log L}} \tag{2}$$

The hyperspectral image contains redundancy and noise information. The singular value decomposition is used to filter the image noise to get the data dictionary in this paper. The specific approach is as follows: First, singular value decomposition is performed on $X = U\Sigma V$, and V is right singular matrix with the size of n*m, U is left

singular matrix with the size of L*m, Σ is semi-positive definite m*m order diagonal matrix whose diagonal elements are singular value. The data dictionary is calculated by the following equation:

$$D = U * \Sigma * V' \tag{3}$$

2.2 Solution of the LRR Model

The rank minimization problem is known to be NP-hard, and the most popular choice is to replace rank with the nuclear norm to make the minimization to be convex relaxation problem. Therefore, the Eq. (1) can be written as follow:

$$\min \|Z\|_* + \lambda \|E\|_{1,2}, \quad s.t. \quad X = DZ + E, \tag{4}$$

where $\|Z\|_*$ is the nuclear norm of Z.

The augmented Lagrange multiplier (ALM) method is quite prevalent in the LRR-related works [12]. In this section, inexact-ALM is extended to handle the proposed LRR problem in (4). By introducing the intermediate variable J, it can be rewritten in the following equivalent form

$$\min \|J\|_* + \lambda \|E\|_{1,2} \quad s.t. \quad X = DZ + E, Z = J \tag{5}$$

With the aid of the Lagrange multipliers Y1 and Y2, (5) can be redefined in the final optimization problem (6)

$$\begin{aligned} \min_{Z,E,J,Y_1,Y_2} & ||Z||_* + \lambda ||E||_{1,2} + < Y_1, X - DZ - E > \\ & + < Y_2, Z - J > + \frac{\mu}{2}(||X - DZ - E||_F^2 + ||Z - J||_F^2) \end{aligned} \tag{6}$$

In (6), each unknown variable of Z, E, J, Y_1, and Y_2 is optimized in iterative fashion that is to successively optimize one variable with the other fixed variables. The solving procedure is summarized in Algorithm 1. In each iteration, the variables J and E are updated by the singular value thresholding operator and the $l_{2,1}$ norm thresholding operator. As for the variable Z, it has a closed solution for each column shown in step 2.2. The steps of algorithm is described as bellow.

Algorithm 1 Solving Problem by Inexact ALM

Input: data matrix X, dictionary D, parameter λ , μ

Output: Z , E

Step 1. Initialize:

$Z = J = 0, E = 0, Y_1 = 0, Y_2 = 0, \mu = 10^{-4}, \mu_{\max} = 10^6, \rho = 1.1, \varepsilon = 10^{-6}$

Step 2. While not converged $X - DZ - E \to 0$ do:

1) Fix the other variables and update J by

$$J = \arg\min \frac{1}{\mu}\|\mathrm{J}\|_* + \frac{1}{2}\| J - (Z + Y_2/\mu)\|_F^2$$

2) Fix the others and update Z by

$$Z = (D^T D + E)^{-1}(D^T X - D^T E + J + \frac{Y_1 D^T}{\mu} - \frac{Y_2}{\mu})$$

3) Fix the others and update E by

$$E = \arg\min \frac{\lambda}{\mu}\| E \|_{2,1} + \frac{1}{2}\|E - (X - DZ + Y_1/\mu)$$

4) Update the multipliers

$$Y_1 = Y_1 + \mu(X - DZ - E), Y_2 = Y_2 + \mu(Z - J)$$

5) Update the parameter μ by $\mu = \min(\mu_{\max}, \rho\mu)$

6) Check the convergence conditions

End while

2.3 The LRBS Algorithm

This section describes the proposed band selection algorithm based on low-rank representation (LRBS). We use the technique of spectral clustering algorithm to compute the eigenvectors of graph Laplacian which weight is composed of the low-rank matrix. The algorithm implementation needs to use K-means to cluster a data set into some subsets. The number of categories k in k-means is obtained by virtual dimensionality (VD) algorithm. Different from the classic K-means, in this paper, we utilized a fixed initial cluster centers instead of the random initial ones.

Assume that $\{\mathbf{b}_l\}_{l=1}^{L}$ is a set of band number of a hyperspectral image cube where $\mathbf{b}_l$ is the l th spectral band represented by a column vector, $\mathbf{b}_l = (b_{l1}, b_{l2}, \cdots, b_{lN})^T$ and $\{b_{li}\}_{i=1}^{N}$ is the set of all N pixels in the l th band image. The implementation steps of the BS method proposed in the paper is listed as following.

Algorithm 2 The LRBS algorithm

Input: All the hyperspectral bands

Output: The selected bands

Initialize: k (obtained by virtual dimensionality(VD) algorithm)

Step 1: Divide all the bands are into N subsets, calculate the euclidean distances between the bands of each set,

Step2: Find the band b_1 with the largest correlation value in the cluster center set.

Step 3: Find the band b_n with the largest value different from b_1 in the left band subsets.

Step 4: k=k+1; if k<p, go to step 2.

Step 5: Using k-means algorithm with the initial center set to cluster k subsets.

Step 6: k=1; calculate the center of the subset.

Step 7: Find the band b_n with the smallest value with the subset center, put b_n in the set $\{b_s\}$.

Step 8: k=k+1; if k<k, go to step 6.

3 Hyperspectral Image Experiments

In this section, we conducted a series of experiments to evaluate the classification performance of the selected bands by different BS methods. The compared algorithms include uniform BS (UBS), minimum estimated abundance covariance (MEAC), multigraph determinantal point process (MDPP), dominant set extraction BS (DSEBS), and the proposed BS method. The experiment platform is listed as follows: the machine operating system used is Windows10, the machine configuration is: CPU frequency 2.50 GHz, running memory 8 GB software environment: MATLAB R2014a. The three hyperspectral datasets we used in the experiment include Purdue University's Indiana Indian Pines image, Salinas Valley image, and the image of University of Pavia.

Table 1 shows the specific band subsets selected using different methods on three datasets. Tables 2, 3, and 4 show the classification accuracy of the bands selected using different methods for hyperspectral classification. It can be seen from Table 2 that the band subsets selected by the LRSB method have higher classification accuracy than other classifications in subsequent classifications such as 1st, 2nd, 6th, 7th, 8th, 9th, and 14th classes in Purdue's data. It can also be seen from Tables 3 and 4 that the band classification selected by the LRBS method generally achieves higher classification accuracy in image classification. Therefore, the experiment shows that the LRBS algorithm provided in this paper can indeed select a subset of frequency bands that perform better.

Table 1. Bands selected by UBS, MEAC, MDPP, DSEBS, LRBS

Data	Methods	Selected bands
Purdue Indian Pines (18 bands)	UBS	1, 14, 27, 40, 53, 66, 79, 92, 105, 118, 131, 144, 157, 170, 183, 196, 209, 220
	MEAC	159, 3, 92, 96, 82, 36, 39, 55, 41, 1, 2, 33, 206, 38, 163, 17, 204, 9
	MDPP	10, 39, 59, 75, 79, 85, 92, 130, 140, 146, 147, 149, 150, 152, 160, 164, 175, 193
	DSEBS	42, 129, 97, 131, 174, 16, 176, 177, 172, 43, 192, 193, 98, 171, 99, 132, 40, 33
	LRBS	72, 35, 31, 147, 167, 128, 112, 65, 108, 109, 137, 189, 155, 162, 170, 182, 200, 210
Salinas (21 bands)	UBS	1, 12, 23, 34, 45, 56, 67, 78, 89, 100, 111, 122, 133, 144, 155, 166, 177, 188, 199, 210, 224
	MEAC	107, 148, 203, 149, 5, 8, 105, 3, 28, 12, 18, 10, 44, 36, 25, 17, 51, 32, 110, 68, 58
	MDPP	1, 8, 11, 22, 27, 28, 50, 57, 58, 65, 90, 99, 105, 119, 123, 134, 142, 157, 175, 191, 204
	DSEBS	99, 101, 16, 119, 177, 112, 44, 46, 120, 47, 131, 175, 196, 121, 17, 102, 174, 180, 187, 135, 42
	LRBS	51, 16, 66, 36, 43, 129, 45, 76, 89, 98, 109, 118, 125, 137, 142, 159, 165, 207, 184, 196, 203
Univ. of Pavia (14 bands)	UBS	1, 9, 17, 25, 33, 41, 49, 57, 65, 73, 81, 89, 97, 103
	MEAC	1, 23, 24, 40, 42, 58, 56, 59, 48, 31, 47, 83, 25, 54
	MDPP	2, 23, 44, 46, 50, 62, 66, 73, 89, 91, 92, 93, 96, 102
	DSEBS	86, 102, 64, 20, 21, 63, 65, 6, 19, 22, 7, 66, 95, 67
	LRBS	70, 14, 20, 27, 1, 34, 47, 51, 62, 64, 68, 91, 99, 94

4 Conclusion

This paper proposed an unsupervised low-rank representation of the band selection method. First, the low-rank representation of the hyperspectral image is solved using an inexact -ALM algorithm to obtain a low-rank coefficient matrix. Each column of the low-rank coefficient is used as a vertex of the graph to perform spectral clustering to determine the initial k-means. The clustering center is clustered to finally obtain a band that satisfies the conditions. Experiments on three HSI datasets show that the frequency band selected by the LRBS algorithm can select a better band subset for image classification.

Table 2. P_D, P_{OA}, P_R calculated from the classification results for Purdue's data

Class	Full Bands	UBS	MEAC	MDPP	DSEBS	LRBS
	P_D	P_D	P_D	P_D	P_D	P_D
1	95.65	95.65	93.48	95.65	95.65	100
2	96.01	97.13	93.07	96.08	96.99	94.47
3	96.99	96.51	96.27	97.35	97.23	96.99
4	98.73	98.73	98.31	99.58	98.31	98.31
5	89.44	90.68	91.51	92.34	93.58	94.62
6	97.12	97.67	97.40	96.71	97.12	98.08
7	100	100	100	100	100	100
8	98.78	98.54	99.16	97.49	97.91	99.16
9	100	100	90.00	100	100	100
10	93.93	91.98	93.31	94.65	93.00	95.37
11	94.70	96.13	94.55	95.48	95.85	95.44
12	95.45	94.94	96.29	96.80	97.30	96.80
13	98.54	98.54	99.02	97.56	96.59	98.54
14	93.52	94.15	94.78	94.70	94.55	96.28
15	90.67	95.60	92.49	96.89	93.52	94.30
16	98.92	98.92	98.92	98.92	98.92	98.92
POA	95.09	95.69	94.91	95.89	95.88	**96.09**
PR	97.61	97.90	97.52	98.00	97.99	**98.09**

Table 3. P_D, P_{OA}, P_R calculated from the classification results for Salinas

Class	Full Bands	UBS	MEAC	MDPP	DSEBS	LRBS
	P_D	P_D	P_D	P_D	P_D	P_D
1	95.52	97.16	97.71	97.76	97.16	97.56
2	98.42	98.85	98.44	97.99	99.17	98.95
3	93.78	95.5	94.03	93.98	95.65	95.85
4	95.62	94.69	94.33	97.49	94.74	95.19
5	96.9	96.45	95.19	95.22	96.9	95.44
6	98.79	98.59	98.56	98.79	98.56	97.83
7	98.63	98.21	98.18	97.99	97.65	97.54
8	96.69	95.81	97.4	95.23	96.11	97.14
9	95.87	95.6	94.74	95.29	95.73	95.92
10	96.67	96.37	96.34	96.46	97.25	96.71
11	97.75	97.85	91.1	97.75	98.31	97.1
12	97.15	96.16	95.54	97.46	97.66	96.63
13	96.51	96.94	93.35	96.4	95.63	96.4
14	95.89	98.14	97.66	97.01	98.04	98.88
15	94	95.27	96.52	95.42	95.25	96.42
16	93.3	96.07	93.86	95.07	95.02	95.52
P_{OA}	96.37	96.49	96.45	96.25	96.63	**96.86**
P_R	98.23	98.29	98.27	98.17	98.36	**98.47**

Table 4. PD, POA, PR calculated from the classification results for University of Pavia

Class	Full Bands	UBS	MEAC	MDPP	DSEBS	LRBS
	P_D	P_D	P_D	P_D	P_D	P_D
1	86.42	87.67	86.44	87.97	87.71	77.61
2	73.34	84.38	83.33	82.14	84.63	88.04
3	79.85	78.9	76.66	76.17	79.22	87.18
4	98.81	97.84	98.88	96.95	97.99	77.1
5	91.49	89.93	91.33	93.32	87.11	98.29
6	89.1	91.35	82.78	87.53	87.44	93.26
7	81.1	83.32	76.26	75.64	76.34	89.62
8	78.46	79.09	79.51	79.83	79.3	82.4
9	77.24	75.86	76.32	74.01	76.17	75.53
P_{OA}	84.32	85.19	83.85	84.33	84.25	**85.82**
P_R	96.76	96.93	96.64	96.76	96.75	**97.07**

Acknowledgments. The work of C. Yu is supported by National Nature Science Foundation of Liaoning Province (20170540095) and Fundamental Research Funds for the Central Universities (3132018196).

The work of C.-I Chang is supported by the Fundamental Research Funds for Central Universities under Grant (3132016331).

References

1. Lu, X., Zhang, W., Li, X.: A hybrid sparsity and distance-based discrimination detector for hyper-spectral images. IEEE Trans. Geosci. Remote Sens. **PP**(99), 1–14 (2018)
2. Song, Meiping, Chang, Chein-I: A theory of recursive orthogonal subspace projection for hyper-spectral imaging. IEEE Trans. Geosci. Remote Sens. **53**(6), 3055–3072 (2015)
3. Song, M., Chen, S.-Y., Li, H.-C., Chen, H.-M., Chen, C.C.-C, Chein-I, C.: Finding virtual signatures for linear spectral mixture analysis. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **8**(6), 2704–2719 (2015)
4. UI Haq, I., Xu, X., Shahzad, A.: Band clustering and selection and decision fusion for target detection in hyperspectral imagery. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Taipei, Taiwan,April 19–24, pp. 110 l–1104 (2009)
5. Chunyan, Yu., Chang, Li-Chien Lee Chein-I, Xue, Bai, Song, Meiping, Chen, Jian: BanSpecified virtual dimensionality for band selection: an orthogonal subspace projection approach. IEEE Trans. Geosc. Remote Sens. **56**(5), 2822–2832 (2018)
6. Sun, W., Jiang, M., Li, W., Liu, Y.: A symmetric sparse representation based bandselction method for Hyperspectral imagery classification. Remote Sens. **8**, 238 (2016)
7. Su, H., Du, Q., Chen, G., et al.: Optimized Hyperspectral band selection using particle swarm optimization[j]. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **7**(6), 2659–2670 (2014)
8. Sumarsono, A., Du, Q.: Low-Rank subspace representation for supervised and unsupervised classification of Hyperspectral imagery. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **PP**(99), 1–8 (2016)

9. Lu, X., Chen, Y., Li, X.: Hierarchical recurrent neural hashing for image retrieval with hierarchical convolutional features. IEEE Trans. Image Proces. **PP**(99), 1 (2017)
10. Wang, Y., Lee, L.C., Xue, B., et al.: A posteriori Hyperspectral anomaly detection for unlabeled classification. IEEE Trans. Geosci. Remote Sens., 1–16 (2018)
11. Wang, X., Liu, F.: weighted low-rank representation-based dimension reduction for Hyperspectral image classification. IEEE Geosci. Remote Sens. Lett. **14**(11), 1938–1942 (2017)
12. Cai, J.-F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM J. Optim. **20**(4)

The Application of Culture Particle Swarm Optimization in Frequency Distribution

Zhikai Zhang[1,2] and Fusheng Dai[1,2(✉)]

[1] Harbin Institute of Technology at Weihai, Weihai 264209, China
dfs7113@126.com
[2] Science and Technology on Communication Networks Laboratory, Shijiazhuang 050081, China

Abstract. With the development of communication field in recent years, the demand of TV, mobile phone GPS, and so on is increasing rapidly, the electromagnetic environment is becoming more and more complex, and the problem of spectrum shortage is becoming more and more prominent. In the military field, the modern warfare has evolved into the electronic warfare, electronic defense, electronic interference, and other commonly used tactics become the main means of warfare, if the spectrum resources are not properly managed to allocate, it will lead to the failure of their own equipment interference, loss of battlefield information, leading to the failure of war. Therefore, it is urgent to solve the problem of how to take rational management and distribution of spectrum resources. For the problem of frequency allocation in fixed frequency system, considering the shortcomings of the previous heuristic algorithms (genetic algorithm, particle swarm algorithm, etc.) which are simply evolving in population space, this paper presents a culture particle swarm optimization algorithm based on the culture algorithm of the two-layer evolutionary model, and proves that the method has better performance by simulation.

Keywords: Wireless communication · Frequency distribution · Heuristic algorithm

1 Introduction

Frequency allocation is a typical NP-complete problem. If we use deterministic algorithms such as exhaustive method, staining method, or sequential finger collocation method, it will take a long time to get the result, and only can be applied to the small-scale network [1].

In order to solve this kind of problem, heuristic algorithm arises. This kind of algorithm is not to obtain exact solution, but to obtain the feasible solution of approximate optimal solution in whole space. Heuristic algorithms include simulated annealing algorithm (SA) [2, 3], genetic algorithm (GA) [4–6], ant colony algorithm (ACO) [7–9], etc., which are widely used in various situations both at home and abroad, and solve a lot of practical problems.

Annealing algorithm was first thought in 1953 by N. Metropolis and others were proposed until 1983 to be successfully applied to solving combinatorial optimization

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 1062–1068, 2019.
https://doi.org/10.1007/978-981-13-6264-4_125

problems. It is derived from the principle of solid annealing. When the solid temperature is relatively high, its internal molecules are in a disordered and fast movement of the state, as the temperature of the molecules will eventually slowly stabilize in an orderly state [10]. Genetic algorithm was first proposed by Holland in 1975. It was borrowed from the biological evolution of the "survival of the fittest" law, through genetic crossover and mutation, generation of evolution, the most viable, that is, the most adaptive living environment of individuals. Ant colony algorithm is a heuristic algorithm first proposed by the Dorigo in the 1992, which is based on the study of ant colony foraging, and the ants in ant colony are connected with "pheromone" as media. Ants in the path of foraging through the left pheromone for other ants to provide information to attract them to this path, so the better the path's pheromone concentration will be higher, the probability of being selected is greater. This paper studies the application of the culture particle swarm algorithm in the frequency allocation, then designs the culture particle swarm optimization algorithm to carry on the frequency assignment to the fixed frequency system with the cultural algorithm and the particle swarm algorithm, and uses MATLAB and some classical algorithms to carry on the simulation contrast.

2 An Overview of Frequency Allocation Algorithms

2.1 Frequency Allocation Target

The frequency allocation problem (FAP) is a combinatorial optimization problem in essence, that is, the frequency distribution of radio equipment in a certain region, which makes the target optimization problem. For FAP, the common distribution target is divided into four categories: minimum interference frequency assignment problem, minimum span frequency assignment problem, minimum blocking probability frequency assignment problem, and minimum frequency assignment problem [11]. In this paper, the frequency allocation problem is to minimize the interference to the target frequency allocation.

The total set of available frequency points is F, and the whole communication network is represented by *G(V, E)*. If there are n subnets in the region, $V = \{v_1, v_2, \ldots, v_n\}$, m links, $E = \{e_1, e_2, \ldots, e_m\}$, where the available frequency points of the subnet v are *f (v)*, the number of frequencies required is *D(v),* each link is assigned a frequency $fi \in F$. Set $x_{vf} = 1$ representation to subnet v assignment frequency f, p_{fg} is a system default value for allocating frequency f to subnet v and assigning frequency g to subnet w.

Minimum interference frequency assignment problem (MI-FAP) in the case of limited frequency resources, according to some interference constraints, how to make the frequency distribution of the links to the minimum interference value. The expressions are as follows:

$$\begin{aligned} &\min \sum_{vw\in E}\sum_{f\in F(v),g\in F(w)} p_{vw}(f,g)\cdot x_{vf}\cdot x_{wg} \\ &s.t.\quad \textstyle\sum_{f\in F(v)} x_{vf} = D(v) \quad \forall v\in V \\ &\qquad x_{vf}\in\{0,1\} \qquad \forall v\in V, f\in F(v) \end{aligned} \tag{1}$$

2.2 An Overview of the Culture Particle Swarm Algorithm

According to the structure of the cultural algorithm, population space can choose any kind of algorithm that directs population evolution, such as genetic algorithm or particle swarm algorithm, considering that particle swarm algorithm is easy to fall into local optimal solution, this paper chooses particle swarm algorithm as population space, and uses the situation knowledge and normative knowledge of cultural algorithm to guide its evolution. The culture particle swarm algorithm (CA-PSO) is proposed, which has better performance through the interaction and influence of particle swarm algorithm and cultural algorithm.

3 Algorithm Process and Simulation

3.1 The Process of the Culture Particle Swarm Algorithm

The population size is N, the knowledge solution scale is $a\% \times N$, the space search dimension is D, the frequency of D device is represented, the frequency range is [fre_{min}, fre_{max}], and the frequency point is fre_{num}, the individual of the K-generation population i is

$$X_i^k = \left\{x_{i1}^k, x_{i2}^k, \ldots, x_{iD}^k\right\}. \tag{2}$$

(1) Initialize the population space, including the position of the particle X_i^1, the velocity of a particle V_i^1. Set maximum iteration times G, inertia factor ω, learning factor c_1 and c_2, and random number r_1 and r_2.
(2) The adaptive values of all individuals are computed and sorted, and the individual optimal solution p_{best} and the global optimal solution p_{gbest} are initialized.
(3) Initialize the reliability space. According to the fitness, the optimal solution of the former $a\%$ is chosen to initialize the situation knowledge, and the initialization of the canonical knowledge is to set the upper and lower boundaries as fre_{min} and fre_{max}, and the corresponding adaptive value is set to $+\infty$;
(4) Determine whether to renew the confidence space. Considering that there is little knowledge experience in evolutionary prophase, the influence on the evolution of population space is small, and the demand of the latter population space is increasing, so the renewal interval of the reliability space depends on the current generation g, namely,

$$affectstep = N_1 + \frac{G - g}{G} * N_2. \tag{3}$$

The constants N_1 and N_2 are used to regulate the renewal interval in the formula.

(5) If the reliability space is to be updated, then the optimal individual of the current generation is compared with the worst individual in the situation knowledge according to the Collator adjusted value, and the standard knowledge is updated according to the formula (4). If no update is required, skip to (6).

$$\begin{aligned}
l_j^{t+1} &= \begin{cases} x_{i,j}^t & if\ x_{i,j}^t \le l_j^t \text{ or } \mathrm{obj}(x_i^t) < L_j^t \\ l_j^t & \text{otherwise} \end{cases} \\
L_j^{t+1} &= \begin{cases} \mathrm{obj}(x_i^t) & if\ x_{i,j}^t \le l_j^t \text{ or } \mathrm{obj}(x_i^t) < L_j^t \\ L_j^t & \text{otherwise} \end{cases} \\
u_j^{t+1} &= \begin{cases} x_{k,j}^t & if\ x_{k,j}^t \ge u_j^t \text{ or } \mathrm{obj}(x_k^t) < U_j^t \\ u_j^t & \text{otherwise} \end{cases} \\
U_j^{t+1} &= \begin{cases} \mathrm{obj}(x_k^t) & if\ x_{k,j}^t \ge u_j^t \text{ or } \mathrm{obj}(x_k^t) < U_j^t \\ U_j^t & \text{otherwise} \end{cases},
\end{aligned} \tag{4}$$

where l_i and L_i, respectively, represent the bottom bounds of the decision variables and their corresponding adaptive values, and the u_i and U_i, respectively, represent the upper boundary of the decision variable i and the corresponding adaptive value.

(6) Use the knowledge of belief space to exert influence on the population space, and update the velocity and position of the particle according to the formula (5) and (6).

$$v_{i,j}^{k+1} = \begin{cases} \omega v_{i,j}^k + c_1 r_1 (p_{b,ij}^k - x_{i,j}^k) + c_2 r_2 (g_b^k - x_{i,j}^k), \\ \quad if\ x_{i,j}^k < l_j^k \text{ and } x_{i,j}^k < p_{g,j}^k, \\ \omega v_{i,j}^k - c_1 r_1 (p_{b,ij}^k - x_{i,j}^k) - c_2 r_2 (g_b^k - x_{i,j}^k), \\ \quad if\ x_{i,j}^k > u_j^k \text{ and } x_{i,j}^k > p_{g,j}^k, \\ \omega v_{i,j}^k \pm c_1 r_1 (p_{b,ij}^k - x_{i,j}^k) \pm c_2 r_2 (g_b^k - x_{i,j}^k), \\ \quad otherwise \end{cases}, \tag{5}$$

$$x_{i,j}^{k+1} = \begin{cases} freMax & if \quad x_{i,j}^k > freMax \\ freMin & if \quad x_{i,j}^k < freMin \\ x_{i,j}^k + x_{i,j}^{k+1} & otherwise \end{cases}. \tag{6}$$

The "±" uses the rand () function to take "+" or "−" at 50% probability.

(7) To calculate the adaptive degree of the new generation and sort order, update the individual optimal solution p_{best} and the global optimal p_{gbest}.

(8) If the current number of iterations $g > G$ or global optimal solution to meet the requirements of our set, we can think that the iteration is complete, otherwise return (4) continue to evolve.

3.2 Simulation

In wireless communication network, the same frequency and adjacent frequency interference is the main interference factor, and the demand of frequency interval is related to the geographical distance between devices. Therefore, in order to preliminarily verify the correctness of the frequency allocation algorithm, we first randomly set N communication devices in a 1×1 region, as shown in Fig. 1, and calculate the distance between any two devices i, j:

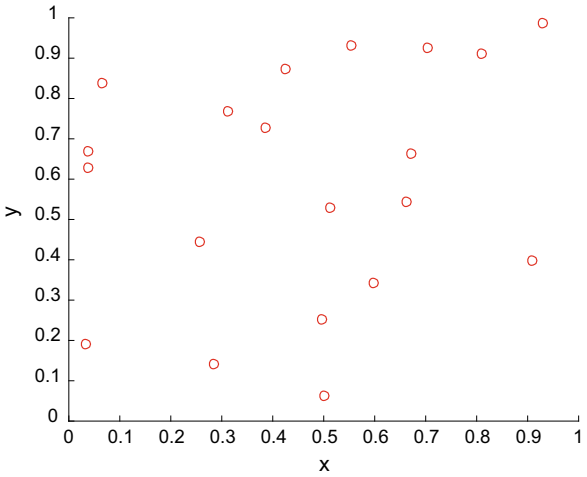


Fig. 1. The distribution of the equipment number is 20

$$d(i,j) = \sqrt{(x(i) - y(i))^2 + (x(j) - y(j))^2}. \tag{7}$$

In this paper, 20 communication devices are set up, the frequency of distribution is 150 MHz–151 MHz, and the simulation is carried out on the frequency interval 0.05 MHz and 0.01 MHz, i.e., 20 frequency points and 100 frequency points, respectively. For the standard of interference, we think that there is no interference when the spacing is large enough, otherwise the closer the distance, the greater the demand for the frequency interval. Therefore, when the equipment spacing is $d < 0.75$ and the frequency interval is $\Delta f < 1 - d$, it is considered to interfere with each other.

The frequency allocation algorithm uses the culture particle swarm optimization algorithm (CA-PSO) and the genetic particle swarm optimization algorithm (GA-PSO), respectively, to simulate the comparison.

First of all, the situation of insufficient spectrum resources simulation, frequency interval of 0.05 MHz. For 20 communication devices, only 20 frequency points are available for distribution, simulation results as shown in Fig. 2.

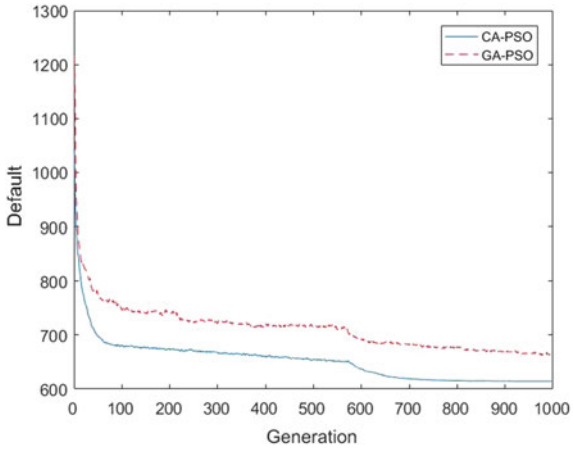


Fig. 2. Frequency distribution when the number of devices is 20 and frequency point is 20

It can be seen from Fig. 2 that when the spectrum resources are not sufficient, the two algorithms are basically convergent after the 1000 generation evolution. Because of the guidance of the belief space, culture particle swarm optimization (CA-PSO) is superior to genetic particle swarm optimization (GA-PSO) in both early and late stages.

Then, consider the situation of sufficient spectrum resources. Frequency interval is set to 0.01 MHz, 100 frequency points are for 20 equipment allocation, and simulation results are as shown in Fig. 3.

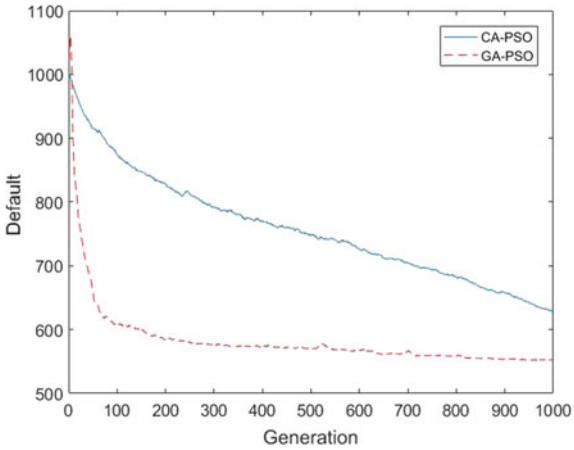


Fig. 3. Frequency distribution when the number of devices is 20 and frequency point is 100

As you can see from Fig. 3, when the spectrum resources are adequate, after 1000 generations of evolution, genetic particle swarm optimization (GA-PSO) has a very early convergence of "premature". And, the culture particle swarm optimization (CA-PSO) algorithm has a long time to update the initial reliability space, so the initial search for the optimal solution is slow, and later, with the knowledge guiding function of the belief space, the results are obviously better than genetic particle swarm optimization.

4 Conclusion

In this paper, we consider that the particle swarm optimization algorithm is easy to fall into the local optimal problem, combined with the cultural algorithm in the population space and the belief space of the two-layer evolutionary model can play a better role in the global population evolution, and put forward the optimal algorithm of culture particle swarm. By comparing the simulation of MATLAB with the genetic particle swarm optimization algorithm, the excellent performance of the culture particle swarm optimization algorithm is validated.

References

1. Ngo, C.Y., Li, V.O.K.: Fixed channel assignment in cellular radio network using a modified genetic algorithm. IEEE Trans. Veh. Tehcnol. **47**(1), 163–172 (1998)
2. Selim, S.Z., Alsultan, K.A.: Simulated annealing algorithm for the clustering problem. Pattern Recogn. **24**(10), 1003–1008 (1991)
3. Bouleimen, K., Lecocq, H.A.: New efficient simulated annealing algorithm for the resource-constrained project scheduling problem and its multiple mode version. Eur. J. Oper. Res. **149** (2), 268–281 (2003)
4. Goldberg, D.E.: Genetic Algorithm in Search (1989)
5. Weuster-Botz, D., Wandrey, C.: Medium optimization by genetic algorithm for continuous production of formate dehydrogenase. J. Dali Univ. **30**(6), 563–571 (2016)
6. Anderson-Cook, C.M.: Practical genetic algorithms. Publ. Am. Stat. Assoc. **100**(471), 1099 (2013)
7. Lu, T., Zhu, J.: Genetic algorithm for Energy-Efficient QoS multicast routing. IEEE Commun. Lett. **17**(1), 31–34 (2013)
8. Zheng, E., Liu, R.: Research on and implementation of ant colony algorithm convergence. Electron. Sci. Technol. **77**(Suppl 1), 107–114 (2013)
9. Dorigo, M., Birattari, M., Stutzle, T.: Ant colony optimization. Comput. Intell. Mag. IEEE **1** (4), 28–39 (2007)
10. Bertsimas, D., Tsitsiklis, J.: Simulated annealing. Stat. Sci. **8**(1), 10–15 (1993)
11. Aardal, Karen I., van Hoesel, Stan P.M., Koster, Arie M.C.A., Mannino, Carlo, Sassano, Antonio: Models and solution techniques for frequency assignment problems. Q. J. Belgian Fr. Italian Oper. Res. Soc. **1**(4), 261–317 (2003)

Spectrum Allocation and Aggregation in Cognitive Radio Satellite Communications

Rong Fu[1], Yuhong Xue[2], Meng Song[1], Shi He[1], Zhen Yang[1], Jinghui Liu[1], Suna Yan[1], and Xin Liu[2(✉)]

[1] Beijing Institute of Control Engineering, China Academy of Space Technology, Beijing 100094, China
37641764@qq.com

[2] School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China
xueyuhong666@163.com, liuxinstarl984@dlut.edu.cn

Abstract. With the rapid development of satellite communications, more and more users and services are flooding into satellite communications networks, causing a shortage of network resources. The most serious of these is the shortage of spectrum resources, which has become one of the bottlenecks to the development of satellite communications. The introduction of cognitive radio technology has greatly eased this conflict. In this paper, focusing on the shortage of KU-band satellite spectrum resources, combining the cutting-edge cognitive radio and spectrum aggregation technology in the field of communications, a KU satellite spectrum allocation strategy based on cognitive radio technology is proposed. By dynamically assigning frequency resources, the use efficiency of spectrum resources is greatly improved and user satisfaction is improved.

Keywords: Cognitive radio · Spectrum allocation · Satellite communication · Graph theory · Spectrum aggregation

1 Introduction

At present, with the development of social and economic, satellite communication has become an indispensable technical equipment, and the functions and reliability of satellite communication systems are also increasingly improved. Satellite communication systems are widely used in many fields such as military, navigation, civil aviation, communications, and maritime search. At present, the KU-band satellite communication system has become the most important part of the ATC satellite communication system [1]. However, the management of satellite transponder resources basically adopts the static allocation method, that is, the spectrum resources on the transponder are often preconfigured according to the task requirements of the system. With the increase in the number of flights, the voice and data capacity for transmission is increasing day by day, and more and more sectors are allocated, resulting in increasingly scarce spectrum resources allocated. The traditional spectrum allocation method can only use continuous spectrum resources. When the user's demand is greater than the continuous spectrum, this part of the cognitive user cannot

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 1069–1076, 2019.
https://doi.org/10.1007/978-981-13-6264-4_126

allocate spectrum to realize communication, so the spectrum is still not fully utilized in the traditional mode. Due to the development of DOFDM, spectrum aggregation technology can realize the integration of discontinuous frequency bands to meet the spectrum requirements of most cognitive users.

2 Basic Concepts of Cognitive Radio and Spectrum Aggregation

2.1 Cognitive Radio Theory

Based on the definition put forward by Joseph Mitola in 2000, it describes the cognitive radio as an integrated agent architecture for software-defined radio to improve intelligence from an intelligent point of view. The physical platform of cognitive radio is similar to that of software radio and is divided into two parts: software and hardware. Based on the cognitive and learning functions, as shown in the platform description, the hardware section includes an antenna, an Radio Frequency (RF), a modulation, an information security module, a baseband processing, and a user interface module. The software part is described by the Radio Knowledge Representation Language (RKRL) [2, 3]. The cognitive radio network model is the cognitive radio network hierarchy proposed by the CRWG by combining CR with the traditional radio model. The model structure is shown in Fig. 1. A complete cognitive wireless network layer includes the physical layer, MAC layer, network layer, and application layer.

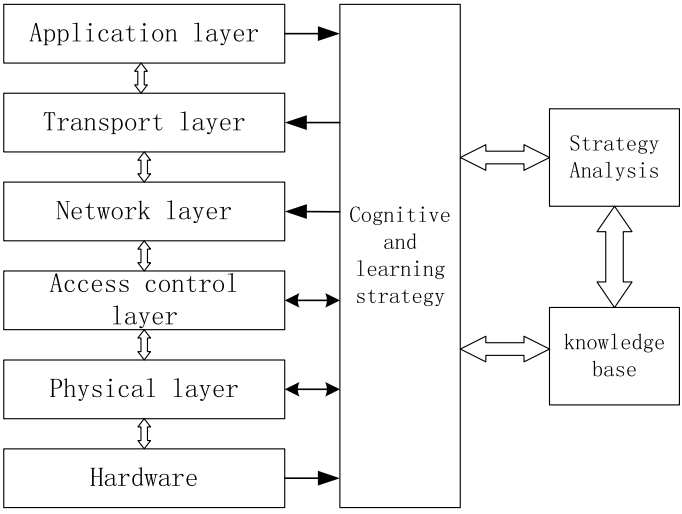


Fig. 1. Cognitive radio network model

The physical layer implements sensing and detection functions, and it recognizes and estimates spectrum, channels, and signal transmissions [4]. Spectrum management

is in the MAC layer, responsible for cognitive users' spectrum allocation. It uses the spectrum to reduce the free resources in the licensed frequency band. At the same time, the MAC layer also needs to implement device coordination and power control, control the transmission power, and other parameters to ensure cooperation and efficient communication between multiple nodes. The network layer needs to consider routes and protocols, communication security issues, and incentive mechanisms. For dynamically changing network topologies, it is necessary to make real-time adjustments such as open licensed bands, and the network layer also needs to solve the security impact of cognitive user access on primary users. The application layer is responsible for building wireless applications based on cognitive radio.

2.2 Basic Concepts of Spectrum Aggregation

Spectrum aggregation technology is studied by many companies and standard organizations around the world. QinetiQ is one of the research institutions trying to use spectrum aggregation technology to maintain broadband services. It uses the principles and implementation methods to study the causes of frequency spectrum fragmentation [5]. As one of the key technologies of LTE-Advanced, spectrum aggregation is attracting attention from more and more companies such as Ericsson, NTT Docomo, ETRI, CMCC, and so on. At the same time, in order to make flexible use of spectrum resources, the Federal Communications Commission (FCC) is also preparing to eliminate spectrum utilization restrictions, which makes spectrum aggregation possible [6]. The spectrum aggregation technology based on the third-generation telecommunication (3G) system is also standardized for LTE-Advanced [7].

3 System Model Architecture

The spectrum allocation problem of cognitive wireless networks is the optimal solution to achieve another goal under the constraint condition. It is a typical combination of optimization problem. The solution to this problem is the graph coloring model, which is a classic NP-complete problem [8]. For different optimization goals, there are greedy algorithms, proportional fairness algorithms, color-sensitive shading algorithms, and distributed K-M algorithms.

In the cognitive wireless network, there are several conditions of authorized users, secondary users, and available frequency bands, which abstract the secondary users into the vertices of the graph. The available frequency bands in the network are identified by different colors. The channel can be marked as "1" for cognitive users and "0" for unavailable channels. The graph coloring model is essentially a 0–1 model. Since users exceeding the interference threshold cannot assign the same frequency band, the associated two vertices on one side of the model cannot be identified as the same color.

In the network environment as shown in Fig. 2, secondary users 1, 2, 3, and 4 and three primary users I II III detect that the frequency bands being used are A, B, and C, and gray coils represent the primary user's communication coverage. The user I uses the frequency C, and there is a cognitive user 2 in the coverage area. Therefore, the

users 1, 3, and 4 can access the frequency band C that the primary user is using. In spectrum allocation, it is not only necessary to consider the primary user, but also need to consider the optimization objective of maximizing spectrum utilization under the interference of the secondary user. This is an optimization problem with constraint conditions and can be solved with a graph coloring model.

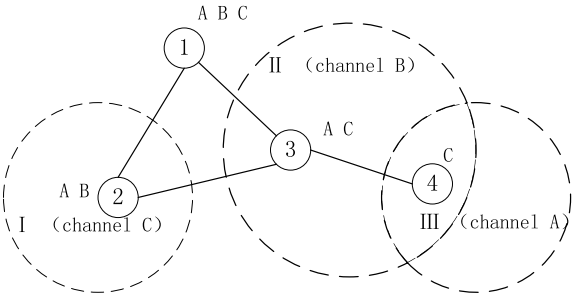


Fig. 2. System model

The definitions of the graph shading model available in this article are listed.

Assume that there are N cognitive users and M available frequency bands in the network after detection, corresponding to the M colors colored in the graph. Available spectrum matrix L, $L = \{l_{nm} | l_{nm} \in \{0,1\}\}_{N\times M}$, is a $N \times M$ order matrix, $l_{nm} = 1$ and indicates that band M is available to user N.

Adjacency matrix E, $E = \{e_{ij} | e_{ij} \in \{0,1\}\}_{N\times N}$, $e_{ij} = 1$ indicates that the use of one frequency band will affect each other and cannot communicate normally.

Allocation matrix A, $A = \{a_{ij} | a_{ij} \in \{0,1\}\}_{N\times M}$, assigns j-band to i-user without interference; set a_{ij} to 1,otherwise set a_{ij} to 0.

Utility matrix B, $B = \{b_{n,m}\}_{N\times M}$, indicates the utility to the system when user n uses band m.

4 Algorithm Description

This algorithm is proposed under the premise that the spectrum aggregation technology can aggregate and distribute the discontinuous frequency bands. The main purpose is to improve the spectrum utilization and user satisfaction, and maximize the system efficiency.

The main steps of the algorithm are shown in Fig. 3, which are described as follows:

Step 1: Establish a subgraph, establish a list of available spectrums for all cognitive users to be assigned, and compare with the demand spectrum number. If the available spectrum number is greater than or equal to the needs of the cognitive user, the sub-picture retains the user for allocation; otherwise, the user is deleted.

Step 2: Select the cognitive users in this distribution process based on the first step, and prioritize the cognitive users with the largest spectrum demand. Therefore,

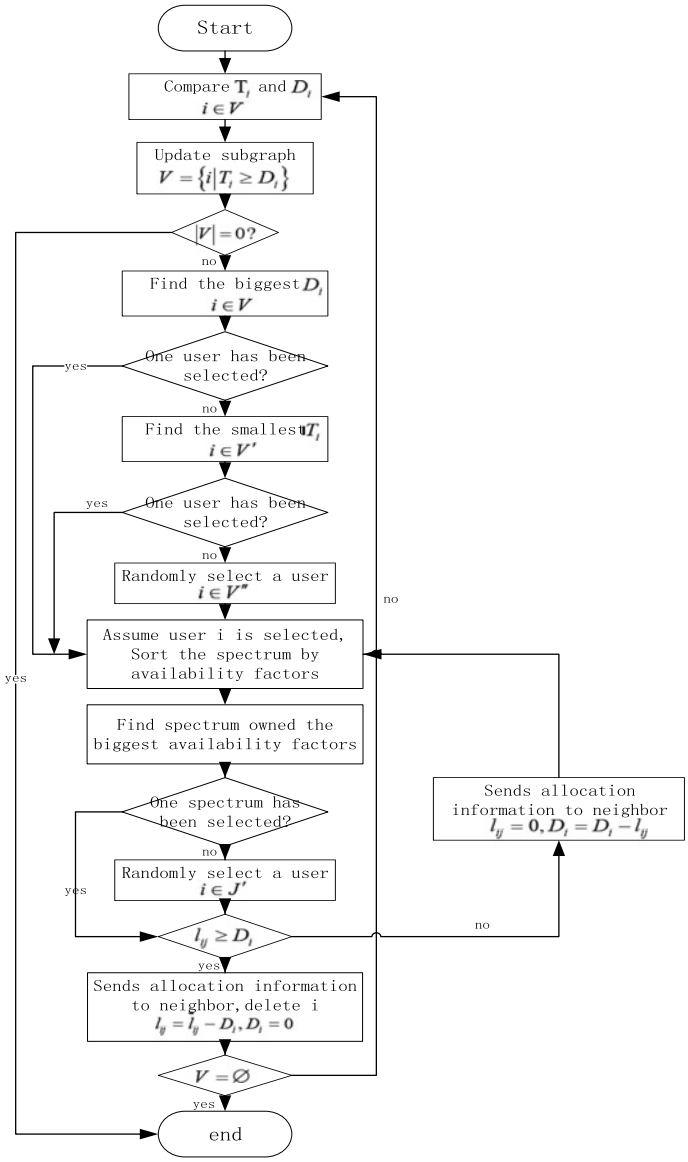


Fig. 3. Algorithm flowchart

arrange the spectrum demand matrix from the largest to the smallest, and select the first one, which has the biggest spectrum requirement. When multiple users have the same requirements, randomly select one user for the next algorithm.

Step 3: The cognitive users who are assigned to the channel choose the appropriate spectrum allocation to meet the user's requirements. The current user has different numbers of interference neighbors in each frequency band. First, the frequency

spectrum is sorted, and the ranking is based on the number of interference nodes in cognitive user's frequency band. Then select the free frequency band with the smallest number of interferences to assign channels to cognitive users. When the number of interferences on the frequency band is the same, the channels are randomly selected for allocation.

Step 4: Send allocation information to neighbor users and update the spectrum matrix. The user is not satisfied to return to the third step. Otherwise, the node is deleted from the subgraph, and the first step is returned. When the subgraph is empty, the algorithm ends.

5 Simulation and Performance Analysis

Simulation parameters are listed in Table 1:

Table 1. Simulation parameters

Parameter	Value
Overall cycle	200
Number of cognitive users N	6
Number of available frequency bands M	3
Demand factors α	1:0.5:5

$D = \{D_i, i = 1 : N\}$ is the cognitive user demand matrix, $T = \{T_i, i = 1 : N\}$ contains the total number of available spectrum units per user,D is randomly generated according to different demand coefficients, and $D \in (0, \alpha * T_i]$. For different demand coefficients, the number of executions of the algorithm flow is K, and the various performance parameters are averaged. The set of users satisfying the demand after spectrum allocation is completed is $S' = \left\{ i | \sum_{j=1}^{M} s_{i,j} \geq D_i \right\}$. Satisfaction is defined as follows:

$$\text{Satisfaction} = \frac{\sum_{i \in S'} D_i}{\sum_{i=1}^{N} D_i} \tag{5.1}$$

Analyze the simulation result graph and compare the traditional list-rendering greedy algorithm and the list-coloring fair algorithm. As shown in Fig. 4 the algorithm proposed in this paper can obtain more user satisfaction under the same demand coefficient, so the algorithm is based on advantage. With the increase of the demand coefficient, the three algorithms tend to be the same in terms of satisfying the user's

demand. This is because most of the users cannot satisfy the available spectrum when the available spectrum is far less than the demand. Since the relationship between user demand and channel bandwidth is taken into account, small demand users do not allocate channels with large bandwidth. Therefore, the proposed algorithm is improved in spectrum utilization compared with other algorithms, as shown in Fig. 5.

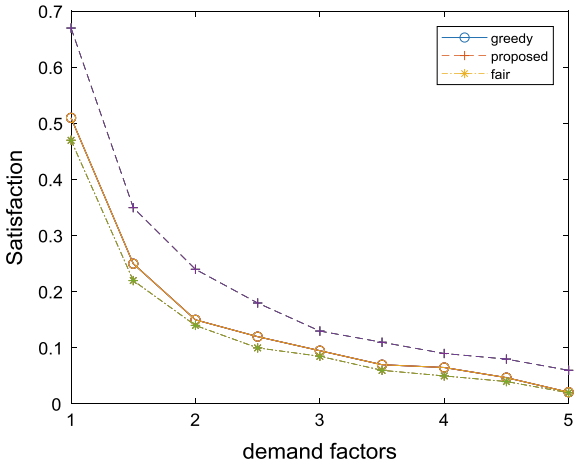


Fig. 4. The relationship between satisfaction and demand coefficient

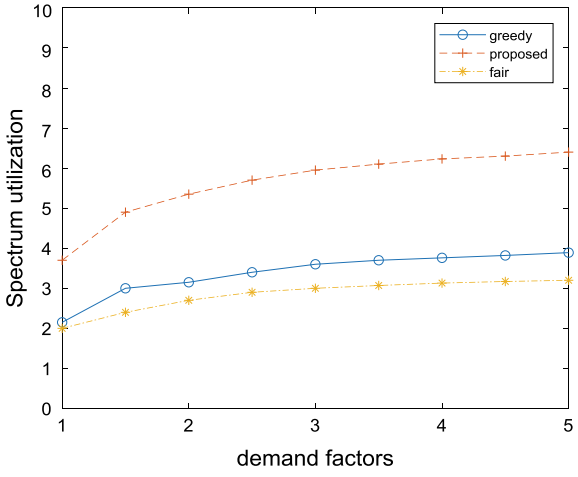


Fig. 5. Relationship between spectrum utilization and demand coefficient

6 Conclusion

The KU-band satellite spectrum allocation scheme based on cognitive radio technology designed in this paper can effectively solve the problem of KU-band satellite spectrum resource shortage. Compared with existing algorithms, this algorithm has advantages in improving user satisfaction and spectrum utilization, but the time is more expensive. In a specific cognitive radio system, suitable algorithms can be selected according to different performance requirements.

Acknowledgements. This paper is supported by the National Natural Science Foundations of China under Grant 61601221; the Joint Foundation of the National Natural Science Foundations of China and the Civil Aviation of China under Grant U1833102.

References

1. Lei, J., Vazquez-Castro, M.A.: Joint power and carrier allocation for the multibeam satellite downlink with individual SINR constraints. In: IEEE International Conference on Communications (ICC), pp. 1–5 (2010)
2. Lagunas, E., Maleki, S., Chatzinotas, S., Soltanalian, M., Perez´-Neira, A.I., Oftersten, B.: Power and rate allocation in cognitive satellite uplink networks. In: IEEE International Conference on Communications (ICC), pp. 1–6 (2016)
3. Zheng, G., Chatzinotas, S., Ottersten, B.: Generic optimization of linear precoding in multibeam satellite systems. IEEE Trans. Wirel. Commun. **11**(6), 2308–2320 (2012)
4. FCC, Facilitating opportunities for flexible, efficient, and reliable spectrum use employing cognitive radio technologies: notice of proposed rule making and order. FCC Document ET Docket No. 03-108 (2003)
5. Zheng, H., Peng, C.: Collaboration and fairness in opportunistic spectrum access. In: Proceedings 40th annual IEEE International Conference on Communications (ICC), pp. 3132–3136 (2005)
6. Huang, J., Berry, R., Honig, M.L.: Auction-based spectrum sharing. ACM Mobile Netw. Appl. (MONET) **11**(3), 405–418 (2006)
7. Tarchi, D., et al.: Technical challenges for cognitive radio application in satellite communications. In: 2014 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), Oulu, pp. 136–142 (2014)
8. Sharma, S.K., Chatzinotas, S., Ottersten, B.: Satellite cognitive communications: Interference modeling and techniques selection. In: 2012 6th Advanced Satellite Multimedia Systems Conference (ASMS) and 12th Signal Processing for Space Communications Workshop (SPSC), Baiona, pp. 111–118 (2012)

Spectrum Sensing Based on Fractional Lower Order Power Spectral Density in Alpha-Stable Noise Environments

Xiaomei Zhu[1], Yongjian Song[1(✉)], Tianjing Wang[2], and Yaping Bao[1]

[1] College of Computer Science and Technology, Nanjing Tech University, Nanjing, JiangSu Province, China
songyongjians@163.com
[2] Department of Information and Computing Science, Nanjing Tech University, Nanjing, JiangSu Province, China

Abstract. In view of the fact that the alpha distribution does not possess two order moment and power spectrum, the detection performance of the traditional detectors, such as energy detector (ED) and power spectral density (PSD) detector, will be degraded or even failed when the background noise be modeled as alpha-stable distribution in CR system. This paper presents a novel spectrum-sensing scheme based on fractional lower order statistics power spectral density (FPSD). The proposed algorithm, combining pseudo-PSD and Fourier transform (FT), calculates the FPSD of the received signal to determine whether primary user (PU) is present or absent. Via the numerous simulations, the performance of the FPSD versus the characteristic exponents α, the moment p, and generalized signal-to-noise (GSNR) of the noise has been studied. Simulations show that the proposed FPSD detector has greater performance than ED and PSD in alpha-stable noise environment. In addition, the new detector, as a blind detector, has high probability detection without the prior knowledge of PU signal and noise.

1 Introduction

Recently, with the increasing demand for wireless devices, the limited spectrum resource of how to be effectively utilized has become a significant problem. In view of this, the cognitive radio (CR) technology has been proposed to utilize efficiently spectral holes and improve the spectrum utilization [1, 2]. In CR network, different from the traditional fixed allocation principle [3], the unlicensed users can access the licensed frequency band to communicate when the primary user (PU) is not occupied on the frequency band. In addition, the unlicensed users must quit immediately when the licensed users will use the spectrum. The spectrum sensing [4–6] that is proposed about those secondary users (SUs) detects whether the presence of PU is one of the most important techniques about a CR system.

The most spectrum-sensing schemes, such as the energy detector (ED) [7–9] and the eigenvalue detector [10–12], are deemed that the background noise is modeled as Gaussian distribution, while in the practical communication environment, the

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 1077–1087, 2019.
https://doi.org/10.1007/978-981-13-6264-4_127

background noise is usually non-Gaussian [13–15], and some spectrum-sensing algorithms based on the ED or the PSD will be declined or even failed [16, 17]. Therefore, some schemes have been introduced to enhance the performance of the detector in non-Gaussian background noise modeled by a symmetric alpha-stable distribution [18–20]. The spectrum-sensing scheme based on the power spectrum density has attracted the attention of researchers, because of its clear physical meaning and easily analysis [21, 22]. Reference [21] has proposed the method based on the difference between the maximum value and minimum value of the power spectrum density (PMMD), which uses the difference between the maximum and minimum values of the received signal power spectrum. PMMD reduces the influence of randomness of the minimum value on the algorithm. However, the alpha-stable distribution has not two order moments and power spectrum. Aiming at this, some algorithms use the fractional lower order statistics (FLOS) in spectrum sensing [23, 24], because the alpha-stable distribution has p-order moments ($p < \alpha$). Therefore, we combine the FLOS and PSD to detect whether the presence of PU is a symmetric alpha-stable distribution noise.

In this paper, we propose new spectrum-sensing scheme-based fractional lower order power spectral density in frequency domain. Both the PU signal and noise are modeled by alpha-stable distribution. The false alarm probability and detection probability are studied in alpha-stable noise environment. A large number of simulations illustrate that the probability detection of the proposed detector is higher than the ED. In addition, the detector also has a superior performance when the prior knowledge of noise, PU signal, is unknown.

The remainder of the paper is arranged as follows. In Sect. 2, the system and alpha-stable distribution noise model are introduced. The spectrum sensing based on fractional lower order power spectral density is provided in Sect. 3. The numerous simulation results of the proposed method and performance analysis are provided in Sect. 4. Finally, Sect. 5 is the conclusion about the performance of the FPSD detector.

2 Problem Formulation

2.1 System Model

Here, we consider a cognitive radio network where there is a primary user (PU), M secondary users (SUs), and a fusion center (FC). In spectrum sensing, each SU detects the presence of the PU signal within a given time interval. The observation of SUs is sent to the FC where the global decision, with regard to the absence or presence of the PU, is made. Hence, the problem of spectrum sensing may be seen as a binary hypotheses test model where H_0 represents the absence of PU and H_1 represents the presence of PU. Under the two hypotheses, the observation vectors $y(n)$ with $n \in \{1, 2, \ldots, N\}$, at the M-th SU, $m \in \{1, 2, \ldots, M\}$, are expressed as

$$\begin{cases} H_0: & y(n) = w_m(n) \\ H_1: & y(n) = s(n) + w_m(n) \end{cases} \tag{1}$$

in which $y(n)$ is baseband signal sample; $w_m(n)$ is the additive background non-Gaussian noise. $s(n)$ is the signal transmitted through the PU. Here, $w_m(n)$ and $s(n)$ are modeled as a symmetric alpha-stable distribution.

2.2 Alpha Distribution Model

In this paper, the noise sequence $w_m(n)$ and the PU signal $s(n)$ are modeled as IID symmetric alpha distribution with the same characteristic exponent α and different dispersion parameters, which are α, γ_w and γ_s respectively. They are as follows:

$$\begin{cases} w_m(n) \sim S\alpha(\beta, \gamma_w, \delta) \\ s(n) \sim S\alpha(\beta, \gamma_s, \delta) \end{cases} \tag{2}$$

In Eq. (2), $\beta = 0$, $\delta = 0$, and $\gamma_w \leq \gamma_s$. The PU signal $s(n)$ and the noise sequence $w_m(n)$ have the same IID symmetric alpha distribution; however, we consider that the dispersion parameter of the noise γ_w is less than that of the PU signal γ_s.

Alpha-stable distribution is expressed by its characteristic function:

$$f(t) = \exp\{j\mu t - \gamma|t|^{\alpha}\} \tag{3}$$

in which $\mu(-\infty < \mu < \infty)$ is the location parameter, and the noise dispersion γ ($\gamma > 0$) is similar to the variance of the Gaussian noise. Figure 1 is the PDF figure of the alpha-stable distribution. It can be viewed that it has a different tail with the various values of the parameter α. When the value of α is equal to 2, it represents that the distribution conforms to Gaussian. What is more, the distribution obeys Cauchy with the value of α equal to 1. Therefore, we select a suitable value of α to fit the non-Gaussian noise in practical cognitive radio networks. The alpha-stable random variable with $\alpha < 2$ implies that there is no finite second moment, so that a majority of traditional spectrum-sensing

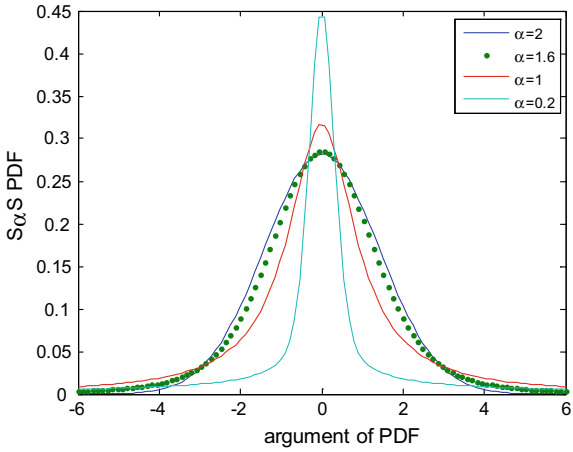


Fig. 1. PDF of alpha-stable distribution with different characteristic exponents α

scheme may not be applied in the alpha-stable distribution. Then, there is only exhibit finite moment for orders less than α, for example, Eq. (4).

$$\begin{cases} E[|X|^p] < \infty, & 0 \le p < \alpha \\ E[|X|^p] = \infty, & p \ge \alpha \end{cases} \tag{4}$$

3 Spectrum Sensing Based on FPSD

3.1 Power Spectral Density Detector

The power spectral density $P_y(k)$ of sampling sequence signal $y(n)$, $0 \le n \le N-1$, with the $y(n)$ can be divided into H groups with the sample length L:$y_h(n)(h = 0, 1, \ldots \mathrm{H}-1, n = 0, 1 \ldots \mathrm{L}-1)$, that can be estimated, i.e.,

$$P_y(k) = \frac{1}{H}\sum_{i=0}^{H-1} P_h(k) = \frac{1}{HL}\sum_{m=0}^{H-1} |Y_h(k)|^2 \tag{5}$$

in which $Y_h(k)$ is the DFT with length L of the received h-th frame signal $y_h(n)$ as follows:

$$Y_h(k) = \sum_{n=0}^{L-1} y_h(n) \mathrm{e}^{-j\frac{2\pi}{N}nk} \tag{6}$$

Some methods based on PSD, such as PMMD, use the difference between the maximum and minimum values of the received signal power spectrum. First, [23] calculate the maximum and minimum values of PSD which are $P_y(\mathrm{k})|_{\max}$ and $P_y(\mathrm{k})|_{\min}$, respectively.

$$P_y(\mathrm{k})|_{\max} = P_y(0) = \frac{1}{H}\sum_{i=0}^{H} |Y_i(0)|^2 \tag{7}$$

$$P_y(\mathrm{k})|_{\min} = P_y(\frac{L}{2}-1) = \frac{1}{H}\sum_{i=0}^{H} \left|Y_i(\frac{L}{2}-1)\right|^2 \tag{8}$$

Then, the sensing statistic can be expressed as

$$T = P_y(0) - P_y(\frac{L}{2}-1) \tag{9}$$

Finally, the detection rules of spectrum sensing can be obtained as follows:

$$\begin{cases} T < \eta, H_0 \\ T > \eta, H_1 \end{cases} \tag{10}$$

where η is the spectrum sensing.

However, the performance of the PMMD is obviously declined or even failed for the non-Gaussian noise environment.

3.2 FPSD Detector

In view of the PSD, the new spectrum-sensing scheme is based mainly on the FPSD of the received signal $y(n)$ by the M-th SU. However, alpha-stable distribution noise has not infinite second-order moment and power spectrum, and we use the FLOS to transform the received signal $y(n)$. Then, the two-hypothesis test model is expressed as

$$\begin{cases} H_0: & z(n) = y(n)^{<p>} = w_m(n)^{<p>} \\ H_1: & z(n) = y(n)^{<p>} = (s(n) + w_m(n))^{<p>} \end{cases} \tag{11}$$

where the symbol $< \cdot >$ is that $z^{<a>} = |z|^a \mathrm{sgn}(z)$ in consideration of the sampling signal $y(n)$ is real number. Similarly, $z(n)$ is divided into H frames with length L. The problem that how FPSD is calculated by discrete Fourier transform (DFT) during the practice should be taken into account. First, $Z_h(k)$ is extracted from $z_h(n)$ for DFT with length L, i.e.,

$$Z_h(k) = \sum_{n=0}^{L-1} z_h(n) \mathrm{e}^{-j\frac{2\pi}{N}nk} = \sum_{n=0}^{L-1} y(n)^{<p>} \mathrm{e}^{-j\frac{2\pi}{N}nk} \tag{12}$$

FPSD is the statistic mean value of $Z_h(k)$ divided by H, i.e.,

$$P_Z(k) = \frac{\sum_{m=0}^{H-1} |Z_h(k)|^2}{HL} \tag{13}$$

where $P_Z(k)$ is fractional lower order pseudo-power spectrum about sample signal $y(n)$.

Under H_0, the sensing statistic is T_1; while in the presence of PU signal, the sensing statistic T_2 is greater than T_1, with the increase of the statistic mean value under H_1. Therefore, according to the characteristics of alpha distribution, the FPSD combine PSD and fractional lower order as proposed. The received signal $y(n)$ is transformed to $z(n)$ by fractional lower order, and $Z_h(k)$ that are extracted from $z_h(n)$ are transmitted to the FC by the M-th SU. Then, the FC based on PSD can obtain the suboptimal sense statistics as follows:

$$T_{FPSD} = \frac{\sum_{m=0}^{M-1}\sum_{h=0}^{H-1}\left|\sum_{n=0}^{L-1} y(n)^{<p>} \mathrm{e}^{-j\frac{2\pi}{N}nk}\right|^2}{HML} \tag{14}$$

It can be seen from Eq. (10) that T_{FPSD} only is related to the size of its DFT, the order p, $0 < p < \alpha/4$, and the sample signal $y_m(n)$, instead of requiring the knowledge of noise power or the non-Gaussian noise PDF, so that the spectrum-sensing scheme under the alpha noise without prior knowledge can detect effectively. Finally, we calculate the value of T_{FPSD}, which is compared with the sensing threshold η for decision that whether PU is present or absent. When $T_{FPSD} > \eta$, it indicates the presence of PU. Otherwise, it means the absence of PU.

Based on the above analysis, determine the appropriate detection threshold, with the given false alarm probability. Then the detection rules of spectrum sensing can be obtained as follows:

$$\begin{cases} T_{FPSD} < \eta, & H_0 \\ T_{FPSD} > \eta, & H_1 \end{cases} \tag{15}$$

3.3 The Main Steps of FPSD

The main steps are summarized as follows:

(1) First, use the FLOS to transform the received signal $y(n)$, and the $z(n)=y(n)^{<p>}$ is obtained.
(2) Second, $z(n)$ is divided into H frames with length L, then $Z_h(k)$ is extracted from $z_h(n)$ for DFT with length L.
(3) Third, T_{FPSD} is the statistic mean value of $Z_h(k)$ divided by H with M-th SU.
(4) Finally, the decision-making process, in which if $T_{FPSD} < \eta$, it shows PU, is absent. If not, the PU is present.

4 Simulation Results and Performance Analysis

In this section, we display some numerous results to investigate the performance of the FPSD spectrum-sensing detector via Monte Carlo simulation in MATLAB. Through this section, we assume that both the PU signal and the non-Gaussian noise follow $S\alpha S$ distribution, and channel is non-fading.

Figure 2 presents the ROC curves for the scheme proposed with different characteristic exponents α ($\alpha = 0.2, 0.8, 1, 1.6, 1.9$) for independent alpha-stable noise; with the length of sample signal sequence $N = 1000$, the simulations of Monte Carlo $MN = 5000$, $GSNR = -2\,\mathrm{dB}$, the number of SU $M = 1$, and the order $p = \alpha/8$. We can observe from the figure that the detection probability P_d improves as α decreases as expected. For example, when $P_f = 0.1$, P_d is obtained as 0.4093 for $\alpha = 1.9$. While $\alpha = 0.2$, the P_d up to 0.8854. It is clear that the FPSD spectrum sensing is feasible.

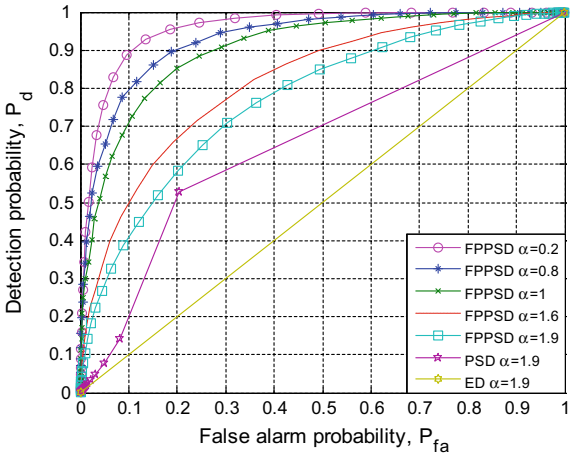


Fig. 2. ROC curve with different characteristic exponents α for independent alpha-stable noise

Especially, Fig. 2 shows that the ED detector and the traditional PSD detector are complete failures in the same conditions for alpha-stable noise environment. That is to say, the performance of FPSD is greater than ED detector and traditional PSD detector for alpha-stable noise.

To illustrate how the characteristic exponents α have influenced the detection performance of FPSD algorithm, the result is simulated in non-fading channel for alpha-stable noise as shown in Fig. 2. With $N = 1000$, $MN = 5000$, $GSNR = -2$ dB, $M = 1$, $P_f = 0.1$, with $p = 0.15/5$, $p = 0.15/8$, $p = 0.15/12$, respectively. With the range of characteristic exponents, α is [0.15, 0.9], recall the FPSD detector. It is shown that the detection probability P_d increases with the characteristic exponents α decrease. For instance, $p = 0.15/8$, when α is equal to 0.9, P_d is only 0.4436, while the value of α is 0.15, the P_d is close to 0.8932, which means that the detection performance of the proposed algorithm improves.

Figure 3 compares the detection probability against different orders p ($p = \alpha/4, \alpha/5, \alpha/8, \alpha/10, \alpha/12$) for fixed $\alpha = 1.6$. Other simulating conditions are $N = 1000$, $MN = 5000$, $GSNR = -2$ dB, $M = 1$, and $P_f = 0.1$. One can see that different orders p lead to the detection probability P_d. For example, when $p = \alpha/4$, the detection probability P_d is approximately 0.5864. However, the P_d is gained as 0.7432 for $p = \alpha/12$. Meanwhile, from the simulation results, we can that the detection probability P_d with optimal order p ($p = \alpha/8$) is greater than $p = \alpha/4$, $p = \alpha/5$, $p = \alpha/10$, and $p = \alpha/12$, respectively. The result depicts that the FPSD spectrum-sensing scheme can get nearly optimal detection performance by selecting a proper order p.

Figure 4 explains the relationship between P_d and p for the proposed scheme. With $N = 1000$, $MN = 5000$, $GSNR = -2$dB, $M = 1$, $P_f = 0.1$, $P_f = 0.1$, and $\alpha = 1.6$. The curves can be obtained with the value of exponent p which is tested from 0.01 to 0.4. As shown in Fig. 5, it can be clearly seen that the optimal p remains around $p = 0.2$. That is to say, the detection performance of FPSD can be improved when the

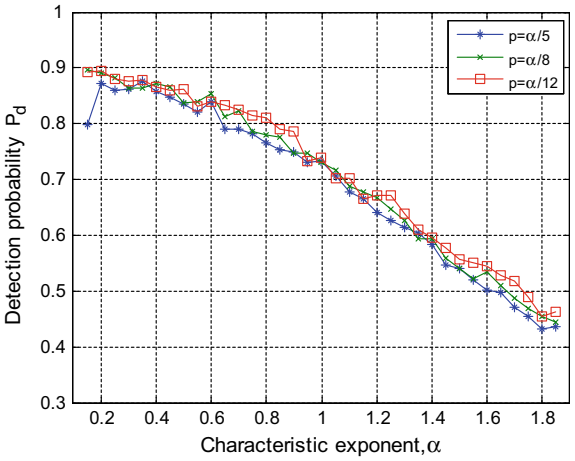


Fig. 3. The detection probability with different characteristic exponents α

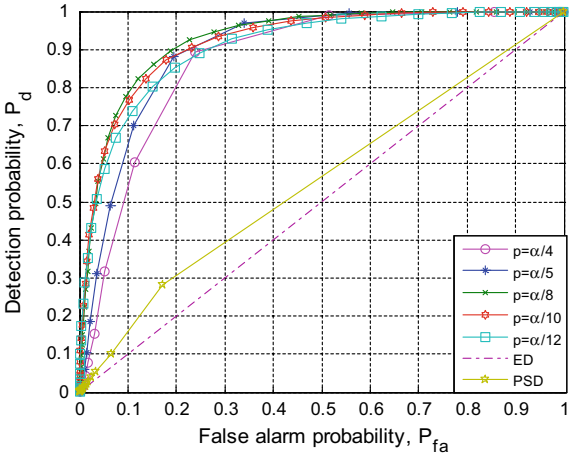


Fig. 4. Comparison of the detection probability against different orders p

order p is close to $\alpha/8$. Particularly, the one can be seen that the ED detector and the traditional PSD detector are complete failures, that is to say, the performance of FPSD is greater than ED detector and traditional PSD detector with the same conditions for alpha-stable noise environment.

To investigate the detection performance of FPSD under different *GSNR*, we simulate the results by selecting the different characteristic exponents α. With $\alpha = 0.9$, $\alpha = 1.2$, and $\alpha = 1.6$, Fig. 6 compares the P_d against different *GSNR*. Other conditions are that $N = 1000$, $MN = 5000$, $M = 1$, $P_f = 0.1$, and $p = 0.9/8$. It is obvious that the detection probability P_d improves as the *GSNR* increases as assumed. Especially, when the $GSNR = -5$ dB and $\alpha = 0.9$, the probability of detection of FPSD is 0.7758. The results verify that the FPSD scheme is effective and feasible.

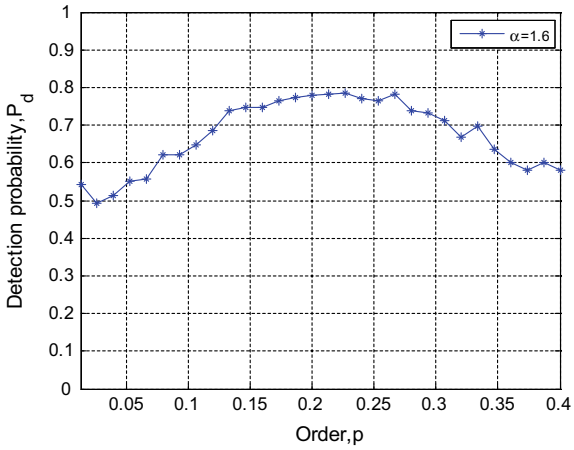


Fig. 5. The relationship among P_d and p for the proposed scheme

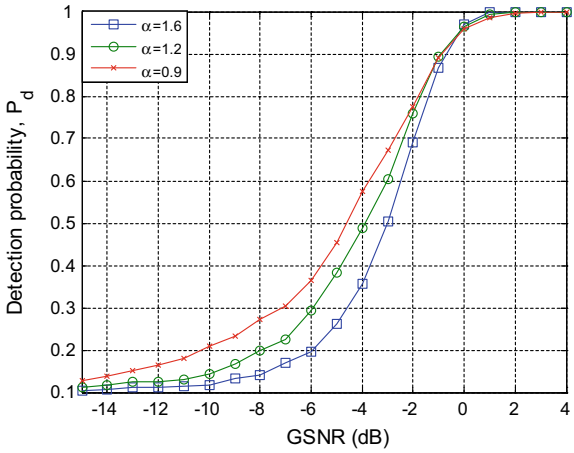


Fig. 6. The comparison of the P_d with different *GSNR*

5 Conclusion

In this paper, we propose a new scheme based on FPSD for an alpha-stable noise. Numerous results prove that the FPSD detector has good performance in several aspects, such as the relationship among the performance detection and characteristic exponents α, moment p, and *GSNR*. In addition, the proposed detector need not have the prior knowledge that noise, channel, and the performance of FPSD are greater than ED detector and PSD detector with the same conditions for alpha-stable noise environment.

Acknowledgment. This work was supported by the National Natural Science Foundation of China under Grant 61501223, 61501224.

References

1. Boulogeorgos, A.A.A., Salameh, H.A.B., Karagiannidis, G.K.: Spectrum sensing in full-duplex cognitive radio networks under hardware imperfections [J]. IEEE Trans. Veh. Technol. **66**(3), 2072–2084 (2017)
2. Ali, A., Hamouda, W.A.: Power efficient wideband spectrum sensing for cognitive radio systems [J]. IEEE Trans. Veh. Technol. **99**, 1 (2017)
3. Zhu, X., Zhu, Y., Bao, Y., et al.: A p th order moment based spectrum sensing for cognitive radio in the presence of independent or weakly correlated laplace noise [J]. Signal Process. **137**(C), 109–123 (2017)
4. Sedighi, S., Taherpour, A., Gazor, S., et al.: Eigenvalue-based multiple antenna spectrum sensing: higher order moments [J]. IEEE Trans. Wirel. Commun. **99**, 1 (2017)
5. Oh, D.C., Lee, Y.H.: Energy detection based spectrum sensing for sensing error minimization in cognitive radio networks [J]. Int. J. Commun. Netw. Inf. Secur. **1**(1) (2009)
6. Sun, M., Zhao, C., Yan, S., et al.: A novel spectrum sensing for cognitive radio networks with noise uncertainty [J]. IEEE Tran. Veh. Technol. **99**, 1 (2017)
7. Chatziantoniou, E., Allen, B., Velisavljevic, V., et al.: Energy detection based spectrum sensing over two-wave with diffuse power fading channels [J]. IEEE Trans. Veh. Technol. **66**(1), 868–874 (2017)
8. Gahane, L., Sharma, P.K., Varshney, N., et al.: An improved energy detector for mobile cognitive users over generalized fading channels [J]. IEEE Trans. Commun. **66**(2), 534–545 (2018)
9. Jin, M., Guo, Q., Li, Y., et al.: Energy detection with random arrival and departure of primary signals: new detector and performance analysis [J]. IEEE Trans. Veh. Technol. **99**, 1 (2017)
10. Kobeissi, H., Nasser, Y., Bazzi, O., et al.: On the performance evaluation of eigenvalue-based spectrum sensing detector for MIMO systems[C]. In: General Assembly and Scientific Symposium, IEEE, pp. 1–4 (2014)
11. Zeng, Y., Liang, Y.C.: Eigenvalue-based spectrum sensing algorithms for cognitive radio [J]. IEEE Trans. Commun. **57**(6), 1784–1793 (2008)
12. Arts, M., Bollig, A., Mathar, R.: Analytical test statistic distributions of the MMME eigenvalue-based detector for spectrum sensing[C]. In: International Symposium on Wireless Communication Systems, IEEE, pp. 496–500 (2015)
13. Moghimi, F, Nasri, A, Schober, R.: Lp-norm spectrum sensing for cognitive radio networks impaired by non-Gaussian noise[C]. In: Global Telecommunications Conference, GLOBECOM, IEEE, pp. 1–6 (2009)
14. Karimzadeh, M., Rabiei, A.M., Olfat, A.: Soft-limited polarity-coincidence-array spectrum sensing in the presence of non-Gaussian noise [J]. IEEE Trans. Veh. Technol. **66**(2), 1418–1427 (2017)
15. Wimalajeewa, T., Varshney, P.K.: Polarity-coincidence-array based spectrum sensing for multiple antenna cognitive radios in the presence of non-Gaussian noise [J]. IEEE Trans. Wirel. Commun. **10**(7), 2362–2371 (2011)
16. Kang, H.G., Song, I., Yoon, S., et al.: A class of spectrum-sensing schemes for cognitive radio under impulsive noise circumstances: structure and performance in nonfading and fading environments [J]. IEEE Trans. Veh. Technol. **59**(9), 4322–4339 (2010)

17. Margoosian, A., Abouei, J., Plataniotis, K.N.: An accurate kernelized energy detection in Gaussian and non-Gaussian/impulsive noises [J]. IEEE Trans. Signal Process. **63**(21), 5621–5636 (2015)
18. Ma, S., Zhao, C., Wang, Y.: fractional low order cyclostationary spectrum sensing based on eigenvalue matrix in alpha-stable distribution noise [C]. In: First International Conference on Pervasive Computing, Signal Processing and Applications, IEEE Computer Society, pp. 500–503 (2010)
19. Chen, R., Wang, J., Lin, R., et al.: Spectrum sensing based on nonparametric autocorrelation in wireless communication systems under alpha stable noise [J]. Mob. Inf. Syst. **2016**(8), 1–6 (2016)
20. Zhu, X., Zhu, W.P., Champagne, B.: Spectrum sensing based on fractional lower order moments for cognitive radios in α-stable distributed noise [J]. Signal Process. **111**(C), 94–105 (2015)
21. Zhao, Z.J., Xi, L.V., Shang, J.N., et al.: The spectrum sensing algorithm using the maximum value and minimum value of the power spectral density [J]. J. Signal Process. (2017)
22. Bomfin, R.C.D.V., Guimarães, D.A., Souza, R.A.A.D.: On the probability of false alarm of the power spectral density split cancellation method [J]. IEEE Wirel. Commun. Lett. **5**(2), 164–167 (2016)
23. Madhavan, N., Vinod, A.P., Madhukumar, A.S., et al.: Spectrum sensing and modulation classification for cognitive radios using cumulants based on fractional lower order statistics [J]. AEUE – Int. J. Electron. Commun. **67**(6), 479–490 (2013)
24. Zhu, X., Jiang, P., Bao, Y., et al.: Spectrum sensing based on fractional lower order moments for cognitive radio under impulsive noise [M]. Electron. Commun. Netw. IV (2015)

A User-Centric Cluster Resource Allocation Scheme for Ultra-Dense Network

Yuzhou Liu(✉) and Wuwen Lai

Beijing Institute of Technology, Beijing, China
{liu_yu_zhou,laiwuwen2013}@bit.edu.cn

Abstract. Ultra-dense network (UDN) is considered as one of the key technologies of 5G. Due to the densification of base stations (BSs) and irregular topology of UDN, BS-centric schemes are usually accompanied by strong inter-cell interference (ICI). In this paper, we propose a user-centric cluster resource allocation scheme (UCRA). In the proposed scheme, we take user experience into consideration and guarantee a desired signal-to-interference-plus-noise-ratio (SINR) for each user equipment (UE) by allocating orthogonal frequency resources to its major interference BSs. When allocating resources, we design an algorithm to seek a balance between system throughput and user fairness. Simulation results show the superiority of our scheme.

Keywords: Ultra-dense network · User-centric · Resource allocation · User experience · User fairness

1 Introduction

Due to the massive application of smart devices, mobile data traffic has been explosively increasing, which puts forward great challenges to the current wireless networks. Ultra-dense network (UDN) is a feasible and effective candidate solution [7]. The basic idea of UDN is to get the access nodes as close as possible to the user equipment (UE). This can be achieved simply by the dense deployment of base stations (BSs) in the hotspots. However, UDN is a typically random network since BSs therein are deployed according to the traffic demands. Consequently, UEs in UDN usually suffer from severe inter-cell interference (ICI).

Generally, methods of interference mitigation are divided into three categories [12]. The first one is to eliminate interference at the receiver. In [9], the authors analyze the performance of zero-forcing receiver and minimum mean-square error receiver with successive interference cancelation under limited backhaul capacity. Notice that the interference signals in this cell have a similar structure to the desired signal because the interference signals are desired signals in their corresponding cell. In light of this, a simultaneous decoding strategy is proposed in [10]. However, the computational complexity of such methods hinders their application.

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 1088–1095, 2019.
https://doi.org/10.1007/978-981-13-6264-4_128

The second method is to adjust the transmission power of the signal dynamically to minimize the ICI. A channel state and interference-aware power allocation scheme based on non-cooperative Cournot model is studied in [3]. Focusing on the multiuser access issue in UDN with non-orthogonal multiple access, an efficient multiuser access scheme is proposed in [11]. However, dynamic adjustment of transmission power implies a complex scheduling algorithm adopted by BSs. Moreover, the receiver complexity increases with the number of UEs.

The third method is to allocate mutually orthogonal radio resources to the potentially interfering UEs so that interference can be avoided. In [1], an adaptive fractional frequency reuse resource allocation scheme is studied. In [4], an orthogonal sub-band assignment based on graph coloring is proposed. Besides, a cluster-based two-stage resource management scheme based on the modified K-means algorithm is proposed in [8] and a coloring-based cluster resource allocation algorithm based on graph theory is studied in [2]. These schemes can be referred to as BS-centric schemes. However, the BS-centric methods are usually with strong ICI due to the densification of BSs and the randomness of UDN topology.

In this paper, we propose a user-centric cluster resource allocation scheme (UCRA). In our scheme, each UE and its cooperating BS constitute a virtual cell. We construct the interference graph based on graph theory, and the coloring algorithm is utilized to categorize UEs into several sets. When allocating resources, we seek a balance between user fairness and system throughput. The rest of this paper is organized as follows. The system model is given in Sect. 2. In Sect. 3, UCRA is stated in detail. Simulation results and analyses are presented in Sect. 4. Finally, conclusions are drawn in Sect. 5.

2 System Model

We focus on the downlink transmission with UCRA. Assuming an orthogonal frequency-division multiple access (OFDMA)-based system and the available bandwidth is divided into several orthogonal resources, namely, resource blocks (RBs).

We consider a channel model incorporating path loss and Rayleigh fading. Suppose UE m is served by BS p, and the propagation gain G_p^m between UE m and its cooperating BS p on unit RB can be given as

$$G_p^m = (d_p^m)^{-\alpha} \cdot h_p^m, \tag{1}$$

where d_p^m and h_p^m are the distance and the exponentially distributed random variable between UE m and BS p, respectively. α denotes path loss exponent.

Similarly, the interference gain G_q^m between UE m and its interfering BS q on unit RB can be expressed as follows:

$$G_q^m = (d_q^m)^{-\alpha} \cdot h_q^m. \tag{2}$$

We assume all BSs have the same transmission power P_0. The $SINR_m$ and the throughput T_m of UE m on unit RB can be expressed as

$$SINR_m = \frac{\sum_{p \subset Z_p} P_0 \cdot G_p^m}{\sum_{q \subset Z_q} P_0 \cdot G_q^m + \sigma^2}, \tag{3}$$

$$T_m = \log_2(1 + SINR_m), \tag{4}$$

where Z_p and Z_q are the set of cooperating BSs and interfering BSs for UE m, respectively. σ^2 is the variance of additional white Gaussian noise (AWGN) on unit RB.

3 A User-Centric Cluster Resource Allocation Scheme

The UCRA includes two parts. In the first part, we construct the virtual cells and system interference graph. The second part is RBs allocation scheme.

3.1 Constructing the Virtual Cells and System Interference Graph

We assume each UE is serviced by its nearest BS, and other BSs are in sleep mode. Besides, each UE and its cooperating BS constitute a virtual cell. The active BSs provide services for those UEs constituting virtual cells with them, and bring interference to others. The BSs in sleep mode do not provide services, nor do they interfere.

Taking user experience into consideration, the SINR of each UE should be above some threshold T_h. We can estimate its SINR based on (3) and compare it with the threshold T_h. If the value is above T_h, there is no severe interference for this UE. Otherwise, the UE suffers from severe ICI. Under this circumstance, it is necessary to find its interfering BSs and remove the largest one. This process goes on until the SINR surpasses the threshold T_h. By then, all the removed BSs are the major interfering BSs for this UE. For interference coordination, the UE is not allowed to share RBs with the UEs served by these BSs. Besides, if one BS serves multiple UEs, these UEs are not allowed to share RBs. To describe the relationship of interference in system, we construct an interference graph, in which each vertex represents a UE. There is an edge between two vertices when these UEs cannot share RBs.

Then, we consider the problem of categorizing UEs into several sets. The UEs in the same set can share RBs with each other. In the interference graph, UEs can share RBs if there is no edge between them. Otherwise, they are not allowed to share RBs. Therefore, we can formulate the UE categorization problem as a graph coloring problem. Given an undirected connected graph and some different colors, use these colors to render the vertices in graph and each vertex has one color. Find a coloring method that makes every two adjacent vertices in the graph have different colors.

Due to the finiteness of orthogonal resources in OFDMA system, we hope that the RB can be shared by more UEs to maximize the utilization of the RB. Therefore, we hope the number of sets as small as possible. The smaller the number of sets, the greater the number of UEs in some sets. We utilize the coloring algorithm, which seeks to solve graph coloring problem with the least number of colors. The vertexes with the same color mean that they can share RBs while vertices with different colors are not. Then, we can categorize UEs into several sets $\boldsymbol{\Psi} = \{\Psi_1, \Psi_2, \cdots, \Psi_C\}$ based on the color, where C is the number of colors needed in coloring algorithm.

3.2 RBs Allocation Scheme

Utilizing the coloring algorithm, we categorize UEs into several sets. Different from average allocation scheme, which allocates resource equally to each set, we consider an optimizing scheme taking account of the balance between system throughput and user fairness.

For system throughput, we consider the current average throughput of virtual cell $T_A^{'}$ as the evaluation indicator.

$$T_A^{'} = \frac{\sum_1^M T_i^{'}}{M} = \frac{\sum_1^M k_i \cdot T_i}{M}, \tag{5}$$

where M is the number of UEs. $\boldsymbol{T}^{'} = \{T_1^{'}, T_2^{'}, \cdots, T_M^{'}\}$ is the current throughput of each virtual cell. $\boldsymbol{k} = \{k_1, k_2, \cdots, k_M\}$ is the number of RB assigned to each cell. $\boldsymbol{T} = \{T_1, T_2, \cdots, T_M\}$ is the throughput of each UE on unit RB.

In our scheme, we seek user fairness on throughput for each UE. Therefore, we consider $(T_A^{'} - T_x^{'})$ as the evaluation indicator for user fairness, where $T_x^{'}$ is the throughput of the current minimum throughput cell (Marked as cell x).

To achieve this, we should maximize $T_A^{'}$ while minimizing $(T_A^{'} - T_x^{'})$. Combining both, the optimization problem can be formulated by

$$\begin{aligned} \max_{\boldsymbol{k}} f(\boldsymbol{k}) &= \max_{\boldsymbol{k}}[T_A^{'} - w \cdot (T_A^{'} - T_x^{'})] \\ &= \max_{\boldsymbol{k}}[(1-w) \cdot \frac{\sum_1^M k_i \cdot T_i}{M} + w \cdot k_x \cdot T_x], \\ s.t. \quad & k_i \geq 0,\ k_i \in Z,\ i = 1, 2, \cdots, M; \\ & \sum_1^C k_j^{'} = K, k_j^{'} \geq 0,\ k_j^{'} \in Z,\ j = 1, 2, \cdots, C. \end{aligned} \tag{6}$$

where w $(0 < w < 1)$ is a weighting factor. $\boldsymbol{k}^{'} = \{k_1^{'}, k_2^{'}, \cdots, k_C^{'}\}$ is the number of RB assigned to each UE set. K is the number of RBs in total.

The optimal solution is hard to find since it is a combinatorial optimization problem. Therefore, we design an algorithm and its main idea is presented as follows.

According to $\boldsymbol{T} = \{T_1, T_2, \cdots, T_M\}$, we can know allocating one RB to which UE set (Marked as set Ψ_Y) can we get the maximum throughput gain. When allocating one RB, we should assign it to set Ψ_Y for system throughput. For user fairness, we should assign it to the set Ψ_X where the cell x is in. When allocating each RB, we compare the schemes of allocating RB to set Ψ_Y and allocating RB to set Ψ_X, and choose the scheme of which the $f(\boldsymbol{k})$ is bigger. The detailed steps about RB allocation scheme are given in Algorithm 1.

Algorithm 1 can be interpreted as a "quantized version" of the gradient ascent algorithm. $f(\boldsymbol{k})$ is the function what we need to maximize. The number of RBs assigned each time corresponds to the learning rate, which determines the length of each step along the gradient. In the process of each gradient rising iteration, we compare scheme Y and scheme X to find which scheme makes $f(\boldsymbol{k})$

Algorithm 1. RBs allocation scheme

Initialization:

$\boldsymbol{\Psi} = \{\Psi_1, \Psi_2, \cdots, \Psi_C\}$: Sets of UEs categorized by coloring algorithm

$\boldsymbol{T} = \{T_1, T_2, \cdots, T_M\}$: Throughput of each UE on unit RB

$\boldsymbol{k} = \{0, 0, \cdots, 0\}$: Number of RB assigned to each cell

K: Number of RBs in total

Preparation:

Making each set $\boldsymbol{\Psi} = \{\Psi_1, \Psi_2, \cdots, \Psi_C\}$ has one RB. Then, $\boldsymbol{k} = \{1, 1, \cdots, 1\}$, Remaining $(K - C)$ RBs

According to $\boldsymbol{T} = \{T_1, T_2, \cdots, T_M\}$,we know allocate one RB to which UE set (Marked as set Ψ_Y) can we get the maximum throughput gain

Main Iteration:

for $i = 1 : (K - C)$ **do**
  Calculating the current throughput of each virtual cell $\boldsymbol{T'} = \{T_1^{'}, T_2^{'}, \cdots, T_M^{'}\}$, $T_j^{'} = k_j \cdot T_j$, $j = 1, 2, \cdots, M$
  Knowing the current minimum throughput cell (Marked as cell x)
  For scheme Y (allocating one RB to set Ψ_Y), we calculate $f_Y(\boldsymbol{k})$
  For scheme X (allocating one RB to set Ψ_X where the cell x is in), we calculate $f_x(\boldsymbol{k})$
  if $f_Y(\boldsymbol{k}) > f_X(\boldsymbol{k})$ **then**
    We execute the scheme Y, allocating one RB to set Ψ_Y
  else
    We execute the scheme X, allocating one RB to set Ψ_X
  end if
end for

increases faster. In other words, we choose the scheme which makes it easier to find the local maximum value of $f(\boldsymbol{k})$. Finally, the local maximum value we find is the global maximum value (up to a quantization error), because the $f(\boldsymbol{k})$ is a convex function.

4 Simulation Results

4.1 Parameters

We consider the Poisson Point Process (PPP) from stochastic geometry theory as the model of UDN [5]. We choose a two-dimensional 1000 square meters area, where BSs and UEs are both randomly and uniformly distributed, following PPP distributions with density parameters of λ_B and λ_U, respectively. In order to evaluate the performance of the UCRA, we compare it with BS-centric methods like AIA scheme [6] in terms of system throughput. Both have similar network scenarios and assumptions. The main parameters used in this paper are presented in Table 1.

Table 1. Simulation parameters

Parameters	Value
BS Density λ_B	500 (/1000 Square Meters)
UE Density λ_U	[50:50:250] (/1000 Square Meters)
BS transmission power P_0	30 dBm
Path loss exponent α	3.5
Variance of AWGN on unit RB σ^2	0.01
Number of RBs in total K	100
RB bandwidth B	180 KHz

4.2 Simulation Results and Analysis

In Fig. 1, the system throughput increases with the number of UEs because the RBs are shared by more and more UEs.

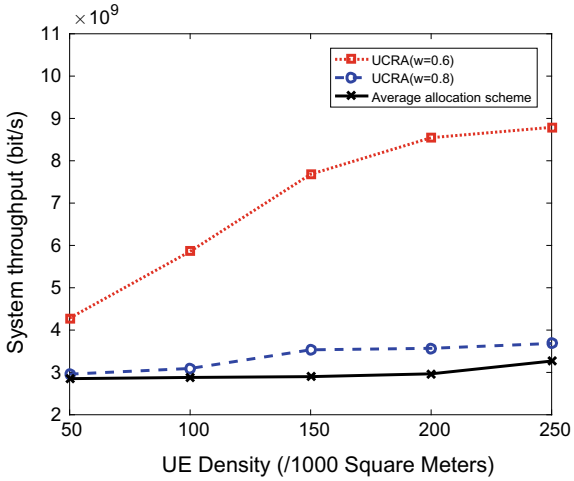


Fig. 1. System throughput vs. Number of UEs

In Fig. 2, the throughput of minimum throughput cell decreases with the number of UEs. On the one hand, more and more BSs are being activated with the increase of UEs, which not only provide service but also bring interference. On the other hand, the number of RBs assigned to each UE will decrease due to the finiteness of orthogonal resources and the increase of UEs.

Besides, we compare our scheme in different values of weighting factor w with the average allocation scheme, which seeks fairness on RB for each UE set. However, our scheme considers fairness on throughput for each UE. When w is 0.6, our scheme aims at system throughput rather than user fairness compared

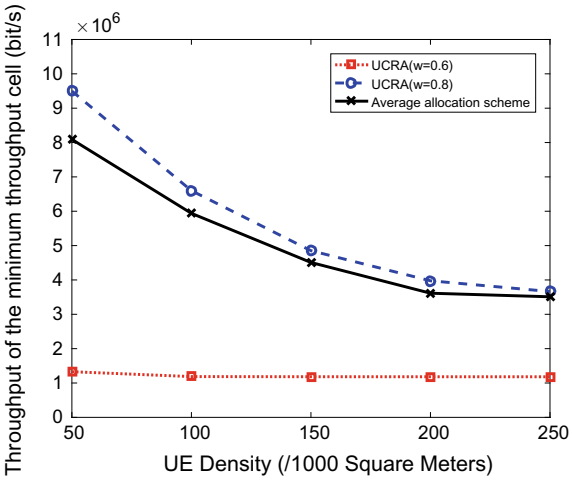


Fig. 2. Throughput of the minimum throughput cell vs. Number of UEs

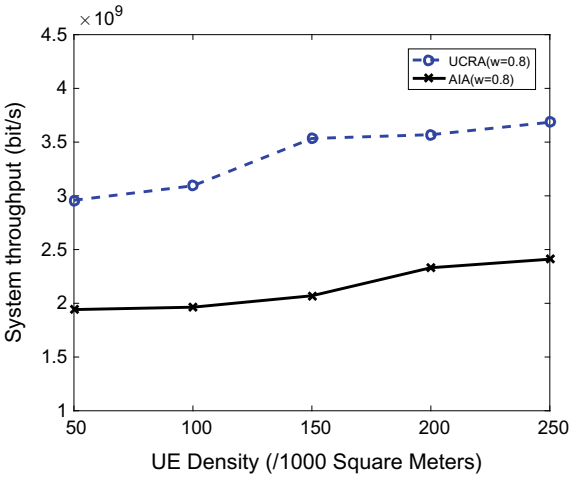


Fig. 3. System throughput vs. Number of UEs

with average allocation scheme. When w is 0.8, our scheme has shown better performance in both ways due to the optimal allocation of RBs.

From Fig. 3, it is clear that our scheme performs better than AIA scheme in system throughput. Due to the densification of BSs and the randomness of UDN topology, BS-centric schemes like AIA usually perform worse in mitigating ICI. As a result, AIA scheme has lower throughput than that of UCRA.

5 Conclusion

In this paper, we propose a user-centric cluster resource allocation scheme (UCRA). In our scheme, each UE and its cooperating BS constitute a virtual cell. We take user experience into consideration and seek a balance between user fairness and system throughput when allocating resources. Simulation results have shown that our scheme outperformed the BS-centric scheme in terms of system throughput.

References

1. Bijan, G., Elvino, S.: Adaptive localized resource allocation with access point coordination in cellular networks. In: IEEE International Conference on Communications, IEEE Press, New York, pp. 1–5 (2011)
2. Zhao, C., Xu, X., Gao, Z., et al.: A coloring-based cluster resource allocation for ultra dense network. In: IEEE International Conference on Signal Processing, Communications and Computing, IEEE Press, New York, pp. 1–5 (2016)
3. Gao, Y., Cheng, L., Zhang X., et al.: Enhanced power allocation scheme in ultra-dense small cell network. In: China Communications, pp. 21–29 (2016)
4. Cao, G., Yang, D., An, R., et al.: An adaptive sub-band allocation scheme for dense femtocell environment. In: Wireless Communications and Networking Conference, IEEE Press, New York, pp. 102–107 (2011)
5. Jeffrey, G., Francois, B., Radha, K.: A tractable approach to coverage and rate in cellular networks. IEEE Trans. Commun. **59**(11), 3122–3134 (2011)
6. Liu, L., Zhu, G., Wu, D.: Interference management based on enhanced fractional frequency reuse in OFDMA networks. In: 7th International Conference on Wireless Communications. Networking and Mobile Computing, IEEE Press, New York, pp. 1–4 (2011)
7. Mahmoud, K., Walaa, H., Amr, Y.: Ultra-dense networks: a survey. IEEE Commun. Surv. Tutor. **18**(4), 2522–2545 (2016)
8. Wei, R., Wang, Y., Zhang, Y.: A two-stage cluster-based resource management scheme in ultra-dense networks. In: IEEE/CIC International Conference on Communications in China, IEEE Press, New York, pp. 738–742 (2014)
9. Wang, S., Yin, H., Wei, G.: Performance analysis of ultra dense network with linear reception and successive interference cancellation under limited backhaul. In: IEEE Conference on Computer Communications Workshops, IEEE Press, New York, pp. 283–287 (2016)
10. Wooseok, N., Dongwoon, B., Jungwon, L., et al.: Advanced interference management for 5G cellular networks. IEEE Commun. Mag. **52**(5), 52–60 (2014)
11. Liu, Y., Li, X., Ji, H. et al.: A multi-user access scheme for throughput enhancement in UDN with NOMA. In: IEEE International Conference on Communications Workshops, IEEE Press, New York, pp. 1364–1369 (2017)
12. Zhou, Y., Liu, L., Du, H., et al.: An overview on intercell interference management in mobile cellular networks: from 2G to 5G. In: IEEE International Conference on Communication Systems, IEEE Press, New York, pp. 217–221 (2014)

Energy-Saving Algorithm Based on Reducing Redundant Uploading Information in Cooperative Spectrum Sensing

Wei Yan, Linjing Zhao(✉), and Sanwen Zhao

State Key Laboratory of ISN, Xidian University, Xi'an, China
wyan_01@stu.xidian.edu.cn, ljzhao@mail.xidian.edu.cn, sanwenzhao@163.com

Abstract. Cooperative spectrum sensing is the key technology that improves the detection performance by using the space diversity gain among cognitive users (CUs). However, it increases the energy consumption. In order to solve the problem of excessive energy consumption, we propose an energy-saving algorithm which can reduce redundant upload information by using the historical sensing results of CUs and the correlation between the sensing results of adjacent time slots. Energy consumption is further reduced by adding a prediction of the channel state. Meanwhile, the proposed algorithms can correct errors caused by the nonideal report channel and improve the detection performance of the system.

Keywords: Cooperative spectrum sensing · Energy saving · Cognitive users · Redundant upload information · Prediction

1 Introduction

With the growth of mobile devices, the demand for wireless spectrum increases exponentially. However, the wireless spectrum is limited due to the scarcity and the dedicated frequency allocation of the standardized wireless system [1]. Cognitive radio (CR) and cognitive radio network (CRN) [2] are popular techniques to analyze the spectrum usage through intelligent detection and transmit data when the spectrum is detected to be idle, thus achieving higher spectrum utilization.

To improve the detection performance, more and more people are interested in cooperative sensing. However, since cooperative sensing increases the information interaction among the cognitive users (CUs) or between the CUs and fusion center (FC), the energy consumption in cooperative sensing is much greater than that of a single user [3]. Meanwhile, according to a survey conducted by the information and communication technology (ICT), more than 70% of the energy is used for spectrum access, while carbon dioxide emissions are very large [4]. Due to the greenhouse effect, people pay more and more attention to energy

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 1096–1104, 2019.
https://doi.org/10.1007/978-981-13-6264-4_129

conservation and emission reduction. Therefore, it is very important to study the energy saving in cooperative sensing.

Authors in [5] reduced the uploading sensing results number of each CU. The algorithm used energy detection techniques and found two optimal decision thresholds. When the detected signal energy value was between two thresholds, it was considered that the sensing result was not useful to data fusion. And CU did not upload the result. This algorithm only reduced the uploading number of the unreliable sensing result. Authors in [6] reduced the sensing results that CUs uploaded to FC. According to the fusion rule, CU only uploaded the result that is useful to the decision. For example, in the "AND" rule, CU only uploaded the result that primary user (PU) did not occupy the channel. This algorithm can greatly reduce the number of uploading sensing results to FC and save energy. Authors in [7] proposed an effective report and accumulation (ERA) scheme that dynamically adjusted the sensing cycle of each CU to improve sensing and reporting efficiency. However, this flexible sensing mechanism could cause some unnecessary collisions. Authors in [8] studied the selection of the CUs to detect the channel and upload the result to the FC. Therefore, the energy consumption was decreased by the reduction of CU number. They proposed two strategies, choosing the user with the highest signal-to-noise ratio and randomly selecting user to upload results. Obviously, this algorithm could significantly reduce the number of CU who detected the spectrum and uploaded the sensing results. In this paper, we focus on how to reduce the uploading information of each CU to save the energy.

The paper is organized as follows. We state the assumptions and describe the system model in Sect. 2. Section 3 presents an energy-saving spectrum-sensing algorithm based on reducing redundant upload information. Section 4 evaluates the proposed algorithm and compares with other schemes. Finally, the paper concludes with Sect. 5.

2 Assumption and System Model

Assume a cooperative spectrum-sensing system as shown in Fig. 1. There are K CUs and an FC in the system, and PU occupies one channel. Similar to Ref. [9], the system is observed in time slots. The duration of each time slot is represented by T. We assume the channel state will not change in a time slot.

Each CU uses the energy detection to sense the channel. In each time slot, CU detects the channel independently and uploads the local sensing result to the FC via the reporting channel [10]. The FC determines the final decision that whether the PU occupies the channel by "AND" rule.

The kth CU energy consumption includes sensing and uploading energy, represented as $e_{s,k}$ and $e_{t,k}$, in a time slot, respectively. If the system is observed in L time slots, the total energy consumed by the kth CU is $E_k = L(e_{s,k} + e_{t,k})$ without any energy-saving operation.

Let $\beta_{k,i} = 1$ indicate that the CU k sends the sensing result in time slot i, and vice versa. The total time slot number of the uploading information by the

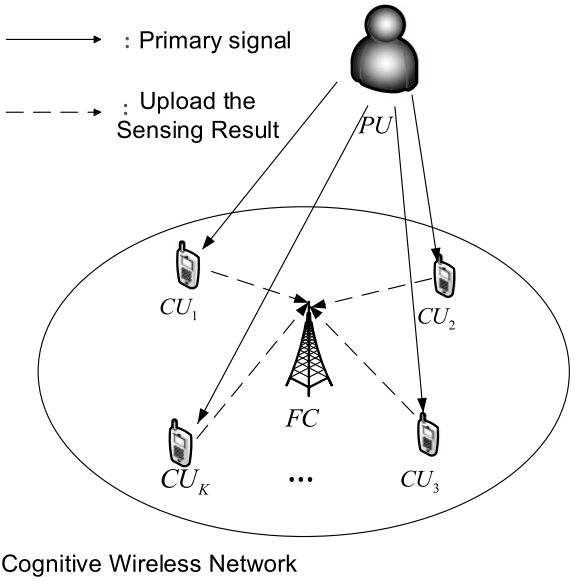


Fig. 1. Cooperative spectrum-sensing scene

CU k is $\sum_{i=1}^{L} \beta_{k,i}$ in L time slots. The total energy consumption of the kth CU is described as

$$E_k = \sum_{i=1}^{L} (e_{s,k} + \beta_{k,i} e_{t,k}). \tag{1}$$

To save energy in a cooperative sensing system, the objective function is defined to minimize the total energy of CU k in L time slots and described as

$$\min E_k = \min \sum_{i=1}^{L} (e_{s,k} + \beta_{k,i} e_{t,k}). \tag{2}$$

Assume that the $\alpha_{k,i}$ represents the local sensing result of CU k in the ith time slot. We will find the solution $\beta_{k,i}$ according to sensing result $\alpha_{k,i}$ to reduce the upload information number $\sum_{i=1}^{L} \beta_{k,i}$, thus saving the energy.

3 The Proposed Energy-Saving Algorithm

In this section, we introduce the spectrum-sensing algorithm based on reducing redundant uploading information in detail. We use the correlation of the sensing results between adjacent time slots to reduce the upload number of sensing results, thereby reducing the energy consumption. Based on the proposed algorithm, we also add a prediction algorithm to reduce more energy consumption.

In order to describe our algorithm clearly, the cooperative spectrum-sensing system is observed in time slots. Each time slot consists of four parts: the CU's local sensing, sensing results uploading, integrating CU sensing results and broadcasting results by FC, and CU communication. Figure 2 illustrates the time slot structure.

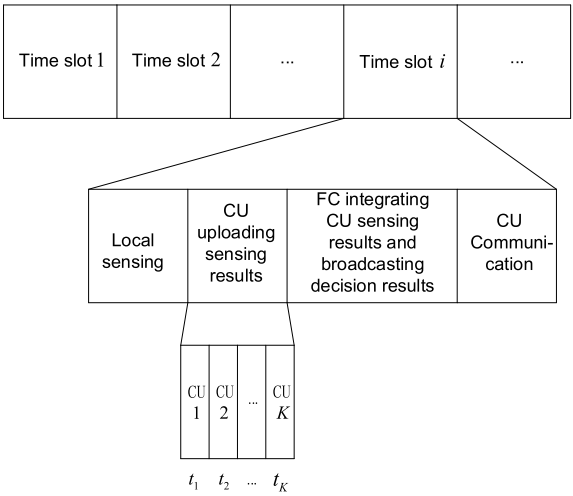


Fig. 2. Time slot structure

3.1 The Proposed Algorithm Based on Reducing Redundant Uploading Information

In this section, we propose the algorithm based on reducing redundant uploading information. First, the CU senses the channel of the PU. Second, CU compares the sensing results of two consecutive time slots and decides whether to send it or not. The FC makes the final decision on the state of the channel and then broadcasts the decision results to the CUs. Specifically, processing in each CU and FC are the main parts of the algorithm. In the following, we will describe our algorithm in detail.

Processing in CU. The first crucial step is the processing in CU. The kth CU senses the channel of the PU in the ith time slot. The sensing result is described as $\alpha_{k,i}$. Then, it will determine whether to transmit the result or not. That means we will find the solution $\beta_{k,i}$ in Eq. (2). We use the correlation of sensing results between adjacent time slots to decide to send the result or not.

The CU k compares the sensing result $\alpha_{k,i}$ of the kth CU in the ith time slot with $\alpha_{k,i-1}$ of the kth CU in the $(i-1)$th time slot. If $\alpha_{k,i} = \alpha_{k,i-1}$, it does not send $\alpha_{k,i}$ to the FC. If $\alpha_{k,i} \neq \alpha_{k,i-1}$, it sends $\alpha_{k,i}$ through the reporting channel. So, we can get

$$\beta_{k,i} = \begin{cases} 0, & \alpha_{k,i} = \alpha_{k,i-1}, \\ 1, & \alpha_{k,i} \neq \alpha_{k,i-1}. \end{cases} \tag{3}$$

Therefore, by this algorithm, the upload information number $\sum_{i=1}^{L} \beta_{k,i} \leq L$ will be reduced. The energy consumption is decreased.

Processing in FC. After each CU reports the sensing result to the FC, the FC first decodes the sensing results of each CU and makes the final decision on whether the channel is occupied or not by the PU.

FC detects whether CU uploads the result or not. Let $\gamma_{k,i}$ be the sensing result of FC received from the kth CU in the ith time slot. If the FC does not receive the data from the kth user in the ith time slot, it means the sensing result of the CU in current time slot is the same as the previous one, $\alpha_{k,i} = \alpha_{k,i-1}$. If FC receives the data from the CU k, it means the channel state detected by the CU k changes.

This rule can also correct the transmission error. If the $\gamma_{k,i} \neq \alpha_{k,i-1}$, it means the FC receives the correct result. If the $\gamma_{k,i} = \alpha_{k,i-1}$, the FC receives error result. For example, $\alpha_{k,i-1} = 1$, $\alpha_{k,i} = 0$, and $\gamma_{k,i}$ should be 0. However, because of the data transmission error, FC receives $\gamma_{k,i} = 1$, same to the data received in $(i-1)$th time slot. FC should correct the error. It reverses the incorrect result and updates $\gamma_{k,i}$. After receiving all the results uploaded by CUs, the FC uses the "AND" rule to make the final decision on whether the PU channel is occupied or not and broadcasts the decision result to the CUs.

The proposed algorithm can save the energy of each CU by reducing the number of uploading, but also can correct the data transmission error. The energy-saving algorithm is described in Algorithm 1.

Algorithm 1. The proposed energy-saving algorithm

Require:
$\alpha_{k,i} = 0$, $i = 1, 2, \ldots, L$,
if $(\alpha_{k,i} = \alpha_{k,i-1})$ **then**
  CU does not upload the result, $\beta_{k,i} = 0$,
else
  CU uploads the result to FC, $\beta_{k,i} = 1$,
  if (FC receives the result) **then**
    if (the result is the same as last time) **then**
      the sensing result is reversed,
    else
      the FC refreshes the storage results,
    end if
  end if
end if
the FC makes a final decision by "AND" rule.

3.2 Reduced Redundant Uploading Information Scheme Based on Prediction

In Sect. 3.1, we propose the energy-saving algorithm based on reducing redundant uploading information. In order to reduce the number of uploading information further, we add a simple prediction method in our algorithm.

The specific rules are similar to the one we proposed in Sect. 3.1. The difference is when the CU sends the sensing result to the FC, the prediction sensing result of the next time slot is also sent. If the sensing result of the next time slot is the same as the predicted one, it will not be sent to the FC. Otherwise, it is uploaded to the FC as usual.

By adding the prediction, the upload number can be reduced further, thereby reducing more energy consumption. The energy-saving algorithm based on prediction is described in Algorithm 2.

Algorithm 2. The proposed energy-saving algorithm based on prediction

Require:
$\alpha_{k,i} = 0$, $i = 1, 2, \ldots, L$,
if ($\alpha_{k,i} = \alpha_{k,i-1}$) **then**
  CU does not upload the result, $\beta_{k,i} = 0$,
else
  CU uploads the result and the prediction result in $(i+1)$th time slot to FC, $\beta_{k,i} = 1$,
  if ($\alpha_{k,i+1}$ = the prediction result) **then**
    CU does not upload the result,
  else
    CU uploads the result to FC,
  end if
  if (FC receives the result) **then**
    if (the result is the same as last time) **then**
      the sensing result is reversed,
    else
      the FC refreshes the storage results,
    end if
  end if
end if
the FC makes a final decision by "AND" rule.

4 Evaluation

4.1 Simulation Parameters

In this section, we simulate the algorithm proposed in this paper. Assume that the channel state follows the Bernoulli distribution model. The time is divided into several time slots, the state of the channel in each slot remains unchanged, and each CU detects the channel state independently. The signal characteristics of the PU follow the Gaussian distribution, and the channel interference is Gaussian white noise with mean 0 and variance 1. Assume that all CUs have the same detection performance, the detection probability is 0.9, and the false alarm probability is 0.1. The energy consumed by the CU k to detect the channel is 150∼200 nJ and to upload the sensing result is 250∼350 nJ [11] in each time slot.

The result is uploaded by BPSK. The bit error rate of the channel is 0.02, the number of simulation slots is 100, and the number of simulations is 10,000. The results received are statistically average.

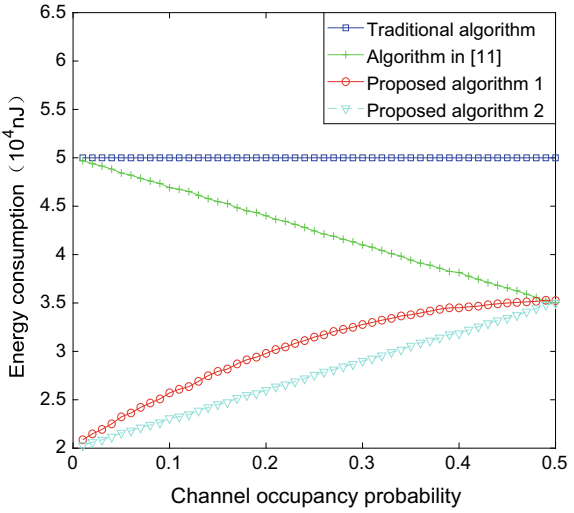


Fig. 3. CU's energy consumption versus channel occupation probabilities

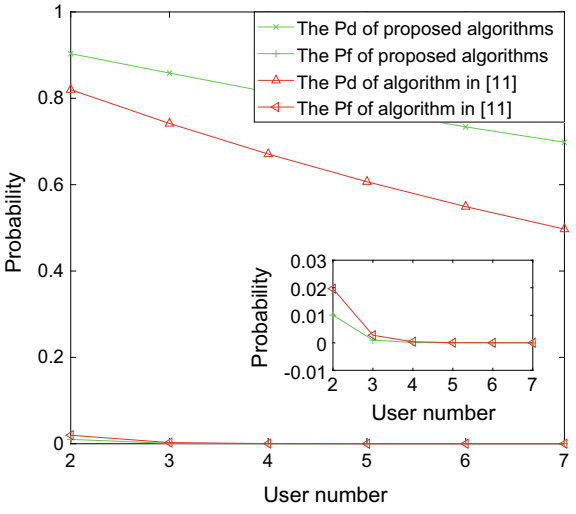


Fig. 4. System detection performance versus the number of cooperative spectrum-sensing users

4.2 Simulation Results

We use "Proposed algorithm 1" and "Proposed algorithm 2" to represent the proposed algorithms described in Sects. 3.1 and 3.2 separately. Performance of proposed algorithms is compared with the traditional algorithm, in which the sensing result will be transmitted in each time slot, and the algorithm in [11], where CU only uploads the sensing result that the channel state is idle.

Figure 3 depicts the CUs' energy consumption versus channel occupation probabilities. In traditional algorithm, the decision whether the CU uploads the result or not is independent with channel occupation probability. Thus, the energy consumption is not changed with the channel occupation probability, and it is a straight line. Compared with traditional algorithm, the algorithm in [11] only uploads the sensing result that the channel is not occupied. However, it only saves parts of energy when the channel occupancy probability is high. In the proposed Algorithm 1, CU only transmits the result when the result is different from the one in the previous time slot. So, it can save more energy, especially when the channel occupancy probability is low. When the idle probability is close to the occupancy probability of the channel, the energy consumption will be increased owing to the number of uploading result increase. However, the maximal energy consumption of our algorithm is still less than the algorithm in [11]. In the proposed Algorithm 2, because of the prediction mechanism, the number of uploading information is decreased further. Based on the simulation results, we conclude that the energy consumption of proposed algorithms is less than the traditional algorithm and the algorithm in [11].

The system detection performance of the algorithm is also evaluated by the system detection probability (P_d) and the system false alarm probability (P_f) as shown in Fig. 4, when the number of system users varies from 2 to 7. Thanks to our proposed prediction mechanism, we can find the error uploaded result and correct it. Thus, the proposed algorithms have higher P_d and lower P_f than the algorithm in [11]. In addition, the users' number influences the system performance. When the users' number rises in the system, the number of misjudgment increases and the detection probability will decrease. Because of the prediction mechanism, the detection probability of our algorithms decreased less. Therefore, the proposed algorithms can improve the detection performance of the system.

5 Conclusion

In this paper, we propose an energy-saving algorithm. It utilizes the correlation between the sensing results of the CU's consecutive time slots and reduces the number of uploading sensing results. We also add a prediction in the proposed algorithm to reduce the energy consumption further. Thanks to the error correction mechanism, our algorithms not only reduce the energy consumption but also enhance the detection performance of the system.

References

1. Sharma, S.K., Bogale, T.E.: Cognitive radio techniques under practical imperfections: a survey. IEEE Commun. Surv. Tutor. **17**(4), 1858–1884 (2015)
2. Pradeep, P.: Research domains for cognitive radio: a survey. In: IT Convergence and Security (ICITCS), 2015 5th International Conference on IEEE, pp. 1–5 (2015)
3. Alfahham, A., Berekovic, M.: Energy efficient cooperative spectrum sensing in cognitive radio sensor network using FPGA: a survey, Open Innovations Association, IEEE, pp. 16–25 (2018)
4. Obaidat, M.S., Anpalagan, A., Wounggang, I.: Handbook of Green Information and Communication Systems. Academic Press (2012)
5. Chen, Y.F.: Analytical performance of collaborative spectrum sensing using censored energy detection. IEEE Trans. Wirel. Commun. **9**(12), 3856–3865 (2010)
6. Sina, M., Geert, L.: Censored truncated sequential spectrum sensing for cognitive radio networks. IEEE J. Sel. Areas Commun. **31**(3), 364–378 (2013)
7. Chang, W., Tai, H.Y., Lee, Y.: c ERA cooperative sensing with differentiated sensing period and retreat scheme in cognitive radio. In: Wireless Communications and Networking Conference, IEEE, pp. 1–6 (2017)
8. Dai, Z.Y., Liu, J., Long, K.P.: Selective-reporting-based cooperative spectrum sensing strategies for cognitive radio networks. IEEE Trans. Veh. Technol. **64**(7), 3043–3055 (2015)
9. Zhang, W.S., Wang, C.X.: Energy-spectral efficiency tradeoff in cognitive radio networks. IEEE Trans. Veh. Technol. **65**(4), 2208–2218 (2016)
10. Pradeep, P.: Kollam, Kerala, research domains of cognitive radio: a survey. In: IT Convergence and Security (ICITCS) 5th International Conference, pp. 1–5 (2015)
11. Maleki, S., Leus, G., Chatzinotas, S., et al.: To AND or to OR on energy-efficient distributed spectrum sensing with combined censoring and sleeping [J]. IEEE Trans. Wirel. Commun. **14**(8), 4508–4521 (2015)

A Spectrum Sensing and Congestion Control Strategy for Distributed Cognitive Radio System

Yi Yang[1], Qinyu Zhang[1(✉)], Ye Wang[1], Takahiro Emoto[2], and Masatake Akutagawa[2]

[1] School of Electronic and Information Engineering, Harbin Institute of Technology (Shenzhen), Shenzhen, People's Republic of China
waiwaisnk25@hotmail.com, zqy@hit.edu.cn, wangye.hitsz@gmail.com
[2] Department of Electrical and Electronic Engineering, Tokushima University, Tokushima, Japan
{emoto,makutaga}@ee.tokushima-u.ac.jp

Abstract. Cooperative spectrum sensing is one of the widely used spectrum sensing methods in cognitive radio systems, which can increase the spectrum sensing quality by leveraging the diversity of multiple secondary users. However, in distributed communication system, there are no effective central controller and cooperative strategy, and most secondary users are likely to overhearing others' sensing results other than contributing to spectrum sensing. This is also called the free-riding attack in distributed cognitive radio system, which will not only make the cooperative spectrum sensing unstable but also causes even more problem in the congestion communication systems. To address the free-riding attack especially in the congested cognitive radio systems, we proposed a cooperative sensing and congestion control strategy with replicator dynamics in evolutionary game theory and a priority system. By using this proposed strategy, rational secondary users have an effective incentive to participate in cooperative sensing, and the associated priority system can eliminate the free-riding attack and congestion problems in communication system, making the cognitive radio system effective and fair. Simulation results show that the average throughput achieved in the proposed cooperative sensing game is higher than the case where secondary users sense the primary user individually without cooperation. And the proposed strategy can also achieve a higher system throughput than the fully cooperative scenario.

Keywords: Distributed cognitive radio networks · Congestion control · Spectrum sensing · Replicator dynamics · Priority system

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 1105–1112, 2019.
https://doi.org/10.1007/978-981-13-6264-4_130

1 Introduction

In order to minimize the primary users' transmission interference caused by secondary users, spectrum sensing is one of essential functions of cognitive radio. Cooperative spectrum sensing is known as a promising method which can greatly improve the spectrum sensing performance [1–6]. The works in [1,5] all focused on the centralized cognitive radios network system that there is a central controller to make a final decision for all the secondary users, and this decision is made to achieve a common goal. In works [6] focused on distributed cognitive radio networks and proposed a cooperative spectrum sensing scheme with selfish secondary users, an evolutionary game is proposed by the authors to model the cooperative spectrum sensing progress, but there is no consideration of secondary users' fairness, the effect of free riding, and all the secondary users can access the licensed spectrum band without congestion problems.

In this paper, we focus on a multiuser distributed cognitive radio system. In order to deal with secondary users' strategy selection, free-riding, and congestion problems, we add a priority system into evolutionary game spectrum sensing strategy. The replicator dynamics in evolutionary game theory will address the strategic uncertainty by exploring different actions, adaptively learning during the strategic interactions, and approaching the best response strategy under changing conditions and environments. The priority system will deal with the congestion and free-riding problems in order to keep a considerable high fairness among secondary users. The proposed spectrum sensing and congestion control strategies could achieve a more steady system than the cognitive radio with only greedy selfish secondary users and a higher average throughput in spectrum sensing than full cooperative spectrum sensing scheme and a better congestion control.

The rest of this paper is organized as follows. In Sect. 2, a distributed cognitive radio system model is shown. In Sect. 3, each secondary users' payoff function is analyzed under different spectrum sensing strategies they have, and we proposed a spectrum sensing and congestion control strategies with priority system in order to reduce the effects of free riding in selfish secondary users. Simulation results are shown in Sect. 4. Finally, Sect. 5 concludes the paper.

2 System Model

In this paper, we focus on the distributed cognitive radio system. Compared to centralized communication system, distributed system is more suitable for the conception of cognitive radio. In this distributed cognitive radio system, we assume that all the secondary users are rational and selfish, secondary users do not serve a common goal of the whole system such as the best detection accuracy and transmission protection for primary user or the biggest throughput of all secondary users. Rational and selfish secondary users in this system only focus on maximizing their own payoff function under certain constraints. These selfish secondary users always take advantage of others by overhearing others

spectrum sensing results if there is any chance. In this paper's cognitive radio communication system, we assume that multiple secondary users will occupy different sub-bands of one primary user when it is absent, and the number of secondary users is more than the number of the sub-carriers of this licensed spectrum. So there are two groups of these secondary users, the first one the accessible ones and the other one the congested ones.

3 Spectrum Sensing and Cognition Control Strategies

3.1 Hypothesis of Spectrum Sensing

As we know, the conception of cognitive radio is to allow the secondary users occupy the licensed spectrum which allocated for primary user. If a secondary user wants to use this licensed spectrum, he must know if the primary user is present or not. Spectrum sensing is designed for this purpose. The primary user's presence status is denoted by hypothesis H_1 and H_0. The received signal is denoted as $r(t)$, and then $r(t)$ can be written as follows:

$$r(t) = \begin{cases} hs(t) + w(t) \\ w(t) \end{cases} \tag{1}$$

In this equation, h is the channel gain between primary user and secondary user, and $s(t)$ is the primary user's transmission signal which is assumed to be an $i.i.d.$ random process with mean 0 and variance δ_w^2. $s(t)$ and $w(t)$ are assumed to be mutually independent.

In this paper, the spectrum sensing method we used is energy detection which is considered as a commonly effective method. Test statistics $T(r)$ is defined as

$$T(r) = \frac{1}{N} \sum_{t=1}^{N} |r(t)|^2 \tag{2}$$

where the number of collected samples is denoted as N.

3.2 Throughput of Secondary User

Under these hypotheses, the throughput of a secondary user is determined by two factors: 1. transmission time. 2. transmission rate. The transmission time for a sensing user is denoted as $T - \delta(N)$, where T is the frame duration, $\delta(N) = \frac{N}{f_s}$ denotes the time spent in sensing, N is the collected samples for energy detection, and f_s is the sampling frequency. So there are two situations here:

1. The primary user is absent.

In these time slots where no false alarm is generated, the average throughput of a secondary user is

$$R_{H_0}(N) = \frac{T - \delta(N)}{T} \times (1 - P_F) C_{H_0} \tag{3}$$

where C_{H_0} is the data transmission rate of the secondary user under H_0.

2. The primary user is present.

In these time slots, the primary user's presence is not detected by the sensing secondary user, and the average throughput of a secondary user is

$$R_{H_1}(N) = \frac{T - \delta(N)}{T} \times (1 - P_D) C_{H_1} \tag{4}$$

where C_{H_0} is the data transmission rate of the secondary user under H_1.

If P_{H_0} represents the probability that the primary user is absent, then in practical cognitive radio system the total throughput of a secondary user can be approximated by

$$\widetilde{R}(N) \approx P_{H_0} R_{H_0} = P_{H_0} \frac{T - \delta(N)}{T} (1 - P_F) C_{H_0} \tag{5}$$

P_F is a decreasing function of N, and as a secondary user tries to reduce $\delta(N)$ in order to have more time to transmit, P_F also will increase at the same time. So there is a trade-off here, and secondary users want to reduce both P_F and N, i.e., keep low false alarm P_F with a smaller N; for this reason, a secondary user intends to cooperate with other secondary users for spectrum sensing in the same licensed spectrum band.

3.3 Spectrum Sensing and Congestion Control Strategies

In this paper, secondary users' spectrum sensing is modeled as a noncooperative game. There are two kinds of strategies $\{C, D\}$ for each secondary user to choose, where strategy C represents contributing to spectrum sensing, and strategy D represents free riding of other secondary users' sensing results. According to Eq. (5), the throughput of a contributed secondary user can be approximated by

$$\widetilde{U}_{C,s_j} = P_{H_0}(1 - \frac{\delta(N)}{|S_C|\,T})(1 - P_F^{S_C}) C_{s_j}, if\ |S_C| \in [1, K] \tag{6}$$

where j is the number jth secondary user chose strategy C, $|S_C|$ is the number of secondary user chose strategy C, and $P_F^{S_C}$ is the false alarm probability calculated by the set of spectrum sensing secondary users S_C. Then, the throughput of a denied secondary user can be approximated by

$$\widetilde{U}_{D,s_i} = P_{H_0}(1 - P_F^{S_C}) C_{s_i}, if\ |S_C| \in [1, K-1] \tag{7}$$

$$\widetilde{U}_{D,s_i} = 0, if\ |S_C| = 0 \tag{8}$$

where i is the number ith secondary user chose strategy D.

In this paper, we chose the majority rule as the decision fusion rules. If the detection probability's threshold given by the primary user is $\overline{P}_D$ for the whole users in the contribution set S_C, then each individual user's target detection probability $\overline{P}_{D,s_j}$ can be calculated by the following equation:

$$\overline{P}_D = \sum_{k=\lceil \frac{1+|S_C|}{2} \rceil}^{S_C} \binom{|S_C|}{k} \overline{P}_{D,s_j}^{k} (1 - \overline{P}_{D,s_j})^{|S_C|-k} \tag{9}$$

and in this contribution set we assumed that each user takes equal responsibility and has the same ability in making the final decision because of fairness concern. Then,

$$P_{F,s_j} = \frac{1}{2} erfc(\sqrt{2\gamma_{s_j} + 1} erf^{-1}(1 - 2\overline{P}_{D,s_j}) + \sqrt{\frac{N}{2\left|S_C\right|}}\gamma_{s_j}) \tag{10}$$

The throughput we discussed above is in the scenario when all the secondary users can access the sub-band at the same time. But if there are congestion problems, the accessibility is the number one thing to be considered with for a secondary user; in another word, no accessibility and no throughput for the secondary user. And there are two congestion scenarios here: 1. The number of secondary users is bigger than the sub-bands' number and the number of cooperative sensing secondary users is smaller than the sub-bands' number; 2. The number of secondary users is bigger than the sub-bands' number and the number of cooperative sensing secondary users is also bigger than the sub-bands' number. The proposed priority system for congestion control works in both scenarios. In scenario number 1, cooperative sensing secondary users have the priority of access and because this kind of users is less than the sub-bands, all the cooperative secondary users can access the sub-bands. The rest of sub-bands will distribute to the uncooperative sensing secondary users depending on their credits' order. In scenario number 2, only the cooperative secondary user outnumbers the sub-bands, so all the noncooperative secondary users are congested for transmission and cooperative secondary user will access the sub-band depending on the credits' order. So the flowchart of this proposed spectrum sensing strategy is shown in Fig. 1, and specific flowcharts of scenario 1 and scenario 2 are shown in Fig. 1a, b.

Figure 2a shows an example of the initial time slot of secondary users' transmission. The blue bar represents the time spent on spectrum sensing, the green bar represents the time spent on low rate transmission, the red bar represents the time spent on high rate transmission. In our proposed spectrum sensing strategy, we combined a priority system into replicator dynamics; so in the initial time slot of secondary users' transmission, there are no credits without contributing to sensing. In other words, if you want to transmit by high transmission rate, you must contribute to sensing in order to consume the permission with the credits earned by sensing.

Figure 2b shows an example of the middle-stage time slot of secondary users' transmission status, and these three bars present the same content as shown in Fig. 2a. In this figure, there are several situations as shown: 1. Secondary users transmit at high transmission rate after contributing to spectrum sensing. 2. Secondary users transmit at high transmission rate without contributing to spectrum sensing. 3. Secondary users transmit at low transmission rate without contributing to spectrum sensing. 4. Secondary users transmit at low transmission rate after contributing to spectrum sensing. Situation NO.1 and Situation NO.3 are easy to understand which can be explained as no pay no gain in single time slot of credits' system. Situation NO.2 means these secondary users

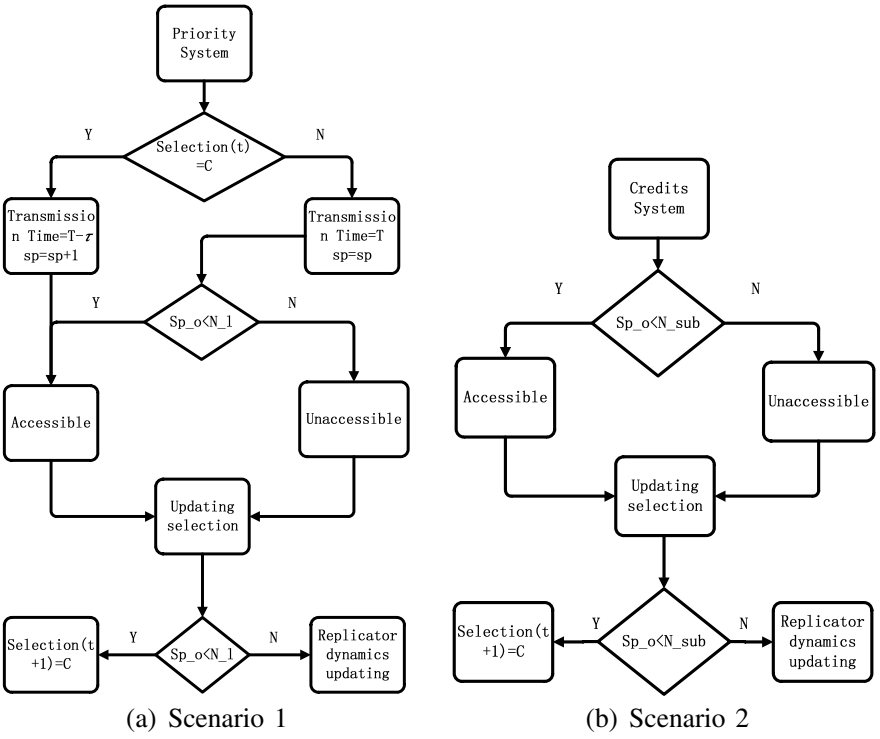


Fig. 1. Flowchart of algorithm

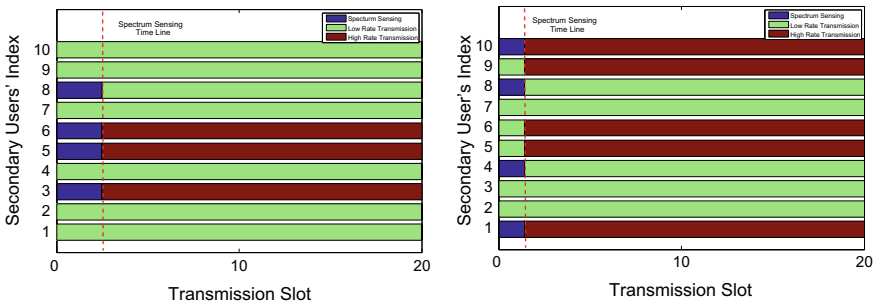
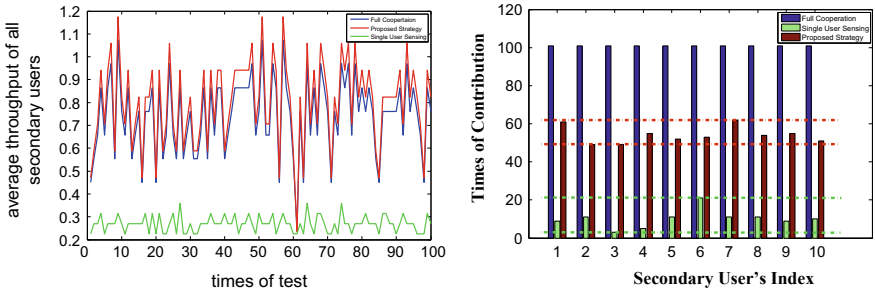


Fig. 2. Secondary users' transmission status

have enough credits before this time slot and they choose to purchase the high transmission rate permission this time slot according to their service category at this time. Situation NO.4 means these secondary want to get some credits without purchasing which will make sure they have enough credits for the coming service no matter they contribute or not at that time slot. Comparing Fig. 2a,



(a) Average Throughput of 100 Times Trial (b) Contribution Times

Fig. 3. Simulation results

b, we can see that the spectrum sensing time are different depending on how many secondary users are contributing to spectrum sensing. The more the users contribute, the less the spectrum sensing spent for each user.

4 Simulation Results

In the simulation, the parameters we used are as follows: We assume that the primary signal is a baseband QPSK-modulated signal with the sampling frequency of $f_s = 1$ MHz and the frame duration of $T = 20$ ms. The probability that the primary user is absent is set as $P_{H_0} = 0.9$, and the required target detection probability $\widehat{P}_D$ is 0.95. The noise is assumed to be a zero-mean CSCG process.

In Fig. 3a, we can see the average secondary users' throughput of the proposed spectrum sensing strategy is mostly bigger than the full cooperation strategy's and they are equal at a certain trial in which the proposed spectrum sensing strategy makes the same decision as the full cooperation sensing strategy. And both proposed spectrum sensing strategy and full cooperation strategy's performance are better than the single-user sensing strategy.

Figure 3b shows secondary users' contribution times of 100 times trial by using these three different spectrum sensing strategies. The bigger the difference between the most contribution times to the lowest contribution times, the more unfair with each secondary user. In simulations, the proposed spectrum sensing strategy has a better fairness performance than the one using the single-user sensing strategy which means cognitive radio network is harder to break down by using the proposed spectrum sensing strategy.

5 Conclusion

In this paper, we proposed a spectrum sensing and congestion control strategies for distributed cognitive radio networks. As the secondary users are selfish and they overhear other's spectrum sensing results, we add a priority system into

replicator dynamics which makes secondary users can try different strategies and learn a better strategy through strategy interactions and ensure secondary users' relative fairness to make the whole system work properly and deal with congestion problem effectively. From the simulation results, the proposed spectrum sensing and congestion control strategies have a better performance of total throughput than fully cooperative strategy which have all secondary users sense at every time slot. Moreover, the proposed spectrum sensing strategy also has a better fairness performance than single-user sensing strategy in order to reduce the effect of free-riding and congestion problems.

Acknowledgment. This work was supported in part by the National Natural Sciences Foundation of China (NSFC) under Grant 61525103, and the Shenzhen Fundamental Research Project under Grant JCYJ20150930150304185.

References

1. Ghasemi, A., Sousa, E.S.: Collaborative specturm sensing in cognitive radio networks. In: Proceedings of IEEE DySPAN 2005, pp. 131–136, Baltimore, MD (2005)
2. Dai, Z.Y., Liu, J.: Selective-reporting-based cooperative spectrum sensing strategies for cognitive radio networks. IEEE Trans. Veh. Technol. **64**(7), 3043–3055 (2015)
3. Chen, X.M., Chen, H.H., Meng, W.X.: Cooperative communications for cognitive radio networks - From theory to applications. IEEE Commun. Surv. Tutor. **16**(3), 1180–1192 (2014)
4. Ganesan, G., Li, Y.: Cooperative spectrum sensing in cognitive radio. IEEE Trans. Wirel. Commun. **6**(6), 2204–2222 (2007)
5. Huang, H., Yuan, C.W.: Cooperative spectrum sensing over generalized fading channels based on energy detection. China Commun. **15**(5), 128–137 (2018)
6. Wang, B., Liu, K.J.R., Clancy, T.C.: Evolutionary cooperative spectrum sensing game: how to collaborate? IEEE Trans. Commun. **58**(3) (2010)

Non-orthogonal Efficient Spectrum Utilization Modulation Scheme Based on Cognitive Radio

Min Jia(✉), Junlong Li, Zhisheng Yin, Qing Guo, and Xuemai Gu

School of Electronics and Information Engineering, Harbin Insitute of Technology, Harbin 150080, Helongjiang, China
jiamin@hit.edu.cn

Abstract. Aiming at the current problem of scarcity and low utilization of radio frequency resources, idle sub-channel judgment based on energy detection is studied on universal software radio peripheral (USRP) platform. An efficient frequency-division multiplexing system is built in the case of evading authorized users during the transmission process and giving the presence of the primary user. Moreover, the cognitive algorithm based on energy detection and blind detection of SEFDM signals is given, the ID-FSD algorithm is introduced, and the performance of the proposed receiver-end ID-FSD detection algorithm is verified. This algorithm can not only improve the frequency utilization rate but also guarantee the BER performance. In addition, this algorithm is applied to realizing the cognitive radio spectrally efficient frequency-division multiplexing system on USRP platform.

Keywords: Cognitive radio · Non-orthogonal · Spectral efficiency · Detection algorithm · USRP

1 Introduction

In recent years, new technologies such as frequency hopping, spread spectrum, OFDM, MIMO, software radio, and cooperative communication have emerged in succession in order to solve the problems of poor electromagnetic environment and lack of spectrum resources. The emergence of these new technologies solved the problems of poor electromagnetic environment and lack of spectrum resources to varying degrees. However, with the continuous expansion of military wireless communication networks, the issue of low spectrum utilization has increasingly become a bottleneck restricting the performance of military wireless communications. Subcarrier spectrum resources are deployed at the minimum interval between orthogonal subcarriers which is guaranteed by OFDM and has high spectrum utilization. However, in the face of faster data transmission rate requirements in the future, and at the same transmission rate, the spectrally efficient frequency-division multiplexing (SEFDM) [1] transmission scheme proposed by Izzat Darwazeh et al. can further compress the subcarrier spacing based on the OFDM spectrum structure and thus the purpose of saving spectrum resources can be achieved.

The basic idea of SEFDM is to save spectrum resources by further compressing subcarrier spacing based on the OFDM spectrum structure. The main idea of literature

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 1113–1120, 2019.
https://doi.org/10.1007/978-981-13-6264-4_131

[2] is to optimize the existing spectrally efficient frequency-division multiplexing (SEFDM) system signal model. The article starts from the aspects of signal duration and frequency spacing, and the spectrum efficiency and energy efficiency of the system are increased through model optimization. Literature [3] is a SEFDM signal reception scheme based on the FFT framework, in which advantage is that it can demodulate SEFDM signals with a large number of subcarriers. The main idea of literature [4] is to combine the advantages of GFDM and SEFDM to achieve the purpose of reducing the out-of-band leakage of the SEFDM signal spectrum.

Recently, some SEFDM experiments have been conducted in wireless communication systems and optical communication systems. In wireless communication systems, multi-carrier aggregation tests [5] are applied in optical communication systems, and a 3.75 Gbit/s 60 GHz millimeter-wave radio frequency fiber (RoF) is adopted for testing [6, 7]. Experiments have concluded that the SEFDM's BER performance is better than OFDM at the same spectral efficiency, while SEFDM has a higher data transmission rate than OFDM at the same bandwidth.

The problem of scarcity of future spectrum resources will be largely solved by the application of the idea of compressing the frequency band to the new generation of ground, no-load, and satellite communications. To sum up, in view of the current scarcity of radio frequency resources and low utilization rate, a high-efficiency frequency-division multiplexing system is set up based on the universal software radio peripheral (USRP) as a hardware platform in the research. At the same time, the cognitive algorithm based on energy detection and blind detection of SEFDM signal is studied, the ID-FSD algorithm is introduced, and the proposed performance and implementation of ID-FSD detection algorithm at the receiver end is verified in the dissertation.

2 Cognitive Scene Description

The cognitive scenario of this document is shown in Fig. 1. Authorized users and authorized base stations communicate within the cell, and spectral resources in specific frequency bands are occupied within the space. At the same time, spectrum holes are detected by unlicensed access points and non-authorized users in real time and idle spectrum are sought over a certain period of time for access [8, 9]. If the authorized user signal is a wideband signal, the presence or absence of the primary user can be distinguished by the non-authorized user by means of energy detection and SEFDM signal blind detection. Under the condition that the bandwidth is limited, part of the SEFDM subcarriers is closed by means of mask mapping, in which the bandwidth of the SEFDM is reduced and the transmission of the primary user is ensured.

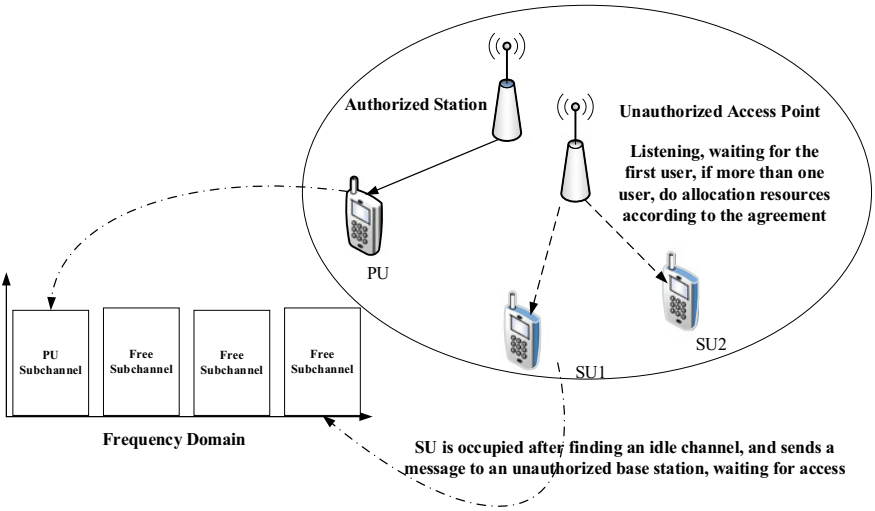


Fig. 1. Unauthorized user access scenarios

3 SEFDM Transmission Model and Cognitive Detection Algorithm

3.1 SEFDM System Transfer Model

The SEFDM system flow diagram is shown in Fig. 2. The compression factor is usually expressed by α, which indicates the degree of compression. For example, when $\alpha = 0.5$, the subcarrier spacing is compressed to 50% (with OFDM as the initial state). The SEFDM spectrum waveforms of the same number of subcarriers for different compression factors are shown in Fig. 2, where $\alpha = 1.0$ corresponds to the OFDM system.

3.2 SEFDM Signal Blind Detection

Consider that the primary user signal cannot be demodulated by unauthorized by the secondary user and cannot be detected using a matched filter. However, assuming that the secondary user knows the bandwidth of the primary user and some basic parameters, the synchronization symbol can be estimated according to the ML maximum likelihood estimation.

After finding the best sub-channel, considering that a cyclic prefix is contained in the SEFDM symbol, the correlation can be used to estimate the symbol timing. The complex baseband vector with an average period of N subcarriers can be regarded as K block and $N + P$ point sampling

$$y = [y(kN + P), \ldots, y((k + 1)(N + P) - 1)] \tag{1}$$

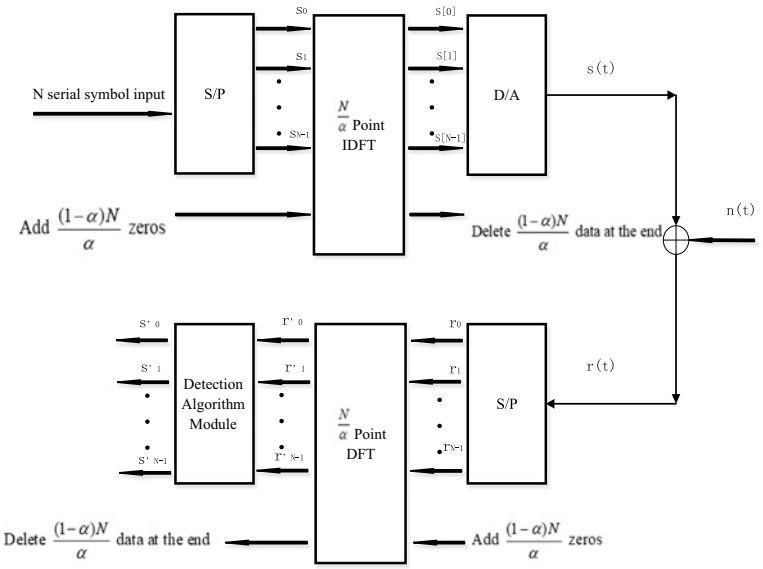


Fig. 2. SEFDM transmission system based on compression factor

where $k = [0, \ldots, K-1]$. The estimation of the correlation delay based on the sliding time window of the cyclic prefix is

$$\begin{aligned}\theta^{opt} &= \max_{\theta \in \{0,\ldots,N+P-1\}} \rho(\theta) \\ &= \max_{\theta \in \{0,\ldots,N+P-1\}} \left\{ \frac{\left| \sum_{k=0}^{K-2} \sum_{j=\theta}^{\theta+P-1} y(k(N+P)+j) y^*(k(N+P)+j+N) \right|}{(K-1)P\sigma_y^2} \right\}\end{aligned} \tag{2}$$

where the variance of the received signal is indicated by σ_y^2; in order to determine whether there is an SEFDM signal, the estimation of the delay should be compared with the non-correlated estimation. Assuming that one channel delay propagates across the entire cyclic prefix, the cyclic prefix sliding window correlation estimation for the non-correlated part can be given by $\rho(\mathrm{mod}(\theta^{opt}+2P, N+P))$. Assuming that the irrelevant part is a Gaussian distribution with mean m_{nc} and variance σ_{nc}^2, then $\eta = m_{nc} + \alpha\sigma_{nc}$ is expressed decision threshold, where α is the standard deviations of $\rho(\theta)$ estimated for the SEFDM signal and the non-SEFDM signal by using the cyclic prefix sliding window. Then according to the threshold, the decision can be drawn as

$$\begin{aligned}&\rho(\theta^{opt}) > m_{nc} + \alpha\sigma_{nc} \text{ SEFDM Signal Exists} \\ &\rho(\theta^{opt}) < m_{nc} + \alpha\sigma_{nc} \text{ SEFDM Signal don't Exists}\end{aligned} \tag{3}$$

3.3 ID-FSD-Based Detection Algorithm

Fixed sphere decoder (FSD) is a breadth-first algorithm based on the sphere-decoding algorithm. Assuming that the data to be detected has a total of M dimensions, the FSD data detection is started from the Mth dimension. The initial detection method is the same as the traditional SD, and only the minimum decision is used starting from the penultimate P-dimension. In order to overcome the problem of FSD optimal path loss, a hybrid algorithm ID-FSD combining iterative detection and fixed sphere decoding is proposed and a better trade-off between performance and complexity can be achieved in this algorithm. The idea of this hybrid algorithm is as follows:

(1) The ID and FSD algorithms run in parallel, and their own detection results are derived for a set of data.
(2) For all symbols, the ID and FSD detection results are used, respectively, to calculate the signal values.
(3) The estimation error of each symbol in the two detection modes is represented by the main diagonal data of the previous calculation result.
(4) The detection result of the detection algorithm with the smaller estimation error calculated in the previous step is selected as the final symbol detection result.

Table 1. ID-FSD algorithm simulation parameters

Detection method	Valid data length	Bandwidth compression factor	FFT points	Decomposition method	ID iterations	SEFDM symbol number	Constellation mapping
ID-FSD	7	$\alpha = 7/8$	8	Cholesky decomposition	10	1000	QPSK
	7	$\alpha = 7/9$	9	Cholesky decomposition	10	1000	QPSK

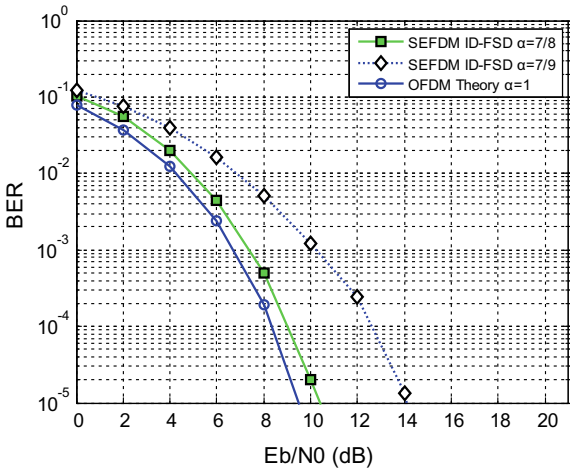


Fig. 3. SEFDM the BER ID-FSD detection mode curve

Thus, a more accurate reference for the FSD algorithm is provided by S_n estimated by the previous algorithm ID, thereby improving the accuracy of the FSD part. The ID-FSD is used for simulation on the AWGN channel. The simulation parameters are shown in Table 1. The simulation results are shown in Fig. 3.

From the simulation results, it can be seen that the ID-FSD performance is excellent, and the FSD algorithm performs is better under the support of the ID algorithm, and the detection complexity of the FSD, especially the limit FSD, is much lower compared to the SD.

4 Implement of USRP Hardware Platform

The basic system parameters in the actual transceiving process are shown in Table 2. QT4 is used in the actual transceiver for IDE development, the UHD driver function is applied, signal processing is performed through the ITpp library [10], Qwt is utilized for plotting the spectrum and constellation diagram, and text encoding is adopted to verify the transmission and reception.

Table 2. Hardware platform specifications

Operating environment	Parameter value
Hardware board	USRP2 (Motherboard) + RFX2400 (Daughterboard)
RF frequency	2.42 GHz
SEFDM points	512 points
Subcarrier mapping	104
Send/receive sampling rate	1 Mbps
Development IDE	QT4

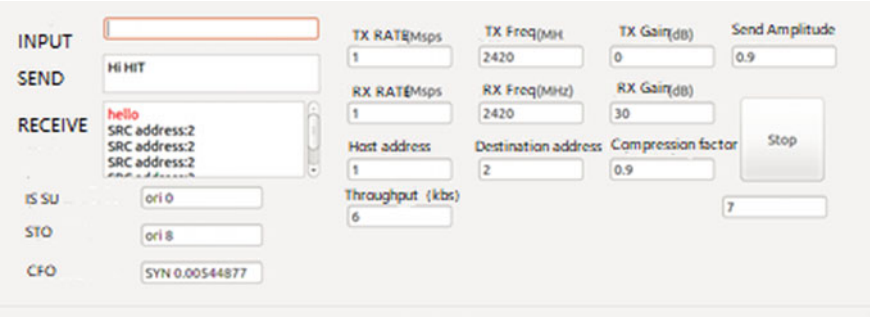


Fig. 4. Sending GUI interface

In Fig. 4, the text that has been sent is indicated by the black font below. In addition, a DC offset component can be seen on the spectrum, which means the simulated authorized users are using. The SEFDM signal is blindly detected in real time indicated

by "IS SU" at the bottom, and time-domain frequency offset estimation is represented by STO and CFO. Since the receiving thread loops and the SEFDM symbol does not blindly detected, so, the maximum correlation window need not be handled.

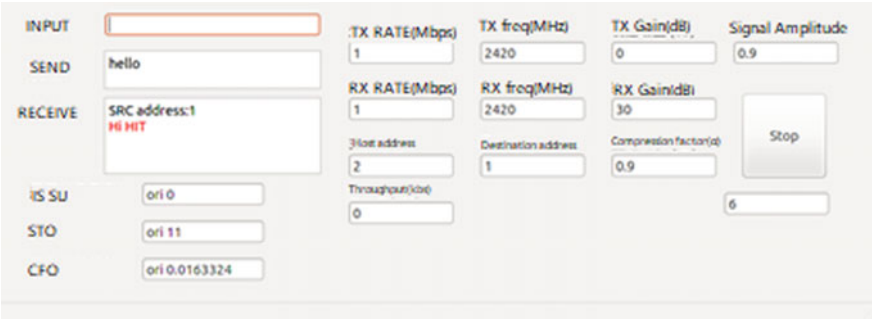


Fig. 5. Text recovered by the receiver

In Fig. 5, the text recovered from the demodulation at the receiving end of another host connected to the USRP is shown in the red part. Compensations are respectively performed, and the phase difference caused by the time domain and the frequency domain is compensated, and then the demodulated text is obtained.

5 Conclusion

In view of the current scarcity of spectrum resources and low utilization rate, the integrated idea of cognitive radio and high-efficiency frequency-division multiplexing technology is introduced in this paper. The ID-FSD detection algorithm at the receiving end is adopted for analysis, and high-efficiency frequency-division multiplexing blind detection algorithm and energy detection algorithm are combined for the simulation of system performance. In addition, ID-FSD algorithm is implemented and validated in the universal software radio peripheral (USRP).

Acknowledgement. This work is supported by National Science Foundations of China (No. 61671183, 61771163 and 91438205).

References

1. Luo, F.L., Zhang, C.: Bandwidth-compressed multicarrier communication: SEFDM, vol. 616. Wiley-IEEE Press (2016)
2. Zavjalov, S.V., Volvenko, S.V., Makarov, S.B.: A method for increasing the spectral and energy efficiency SEFDM signals. IEEE Commun. Lett. **20**(12), 2382–2385 (2016)
3. Rashich, A., Kislitsyn, A., Fadeev, D., et al.: FFT-based trellis receiver for SEFDM signals. In: 2016 IEEE Global Communications Conference (GLOBECOM), pp. 1–6 (2016)

4. Xu, T., Darwazeh, I.: Nyquist-SEFDM: pulse shaped multicarrier communication with sub-carrier spacing below the symbol rate. In: International Symposium on Communication Systems, Networks and Digital Signal Processing (2016)
5. Xu, T., Darwazeh, I.: Bandwidth compressed carrier aggregation. In: IEEE International Conference on Communication Workshop, pp. 1107–1112. IEEE (2015)
6. Mikroulis, S., Xu, T., Darwazeh, I.: Practical demonstration of spectrally efficient FDM millimeter-wave radio over fiber systems for 5G cellular networking. In: SPIE Photonics West, Broadband Access Communication Technologies X (2016)
7. Xu, T., Mikroulis, S., Mitchell, J.E., et al.: Bandwidth compressed waveform for 60-GHz millimeter-wave radio over fiber experiment. J. Lightwave Technol. **34**(14), 3458–3465 (2016)
8. FCC.: Spectrum Policy Task Force Report, in et. Docket no. 02–155. Technical Report (2002)
9. Barbero, L.G., Thompson, J.S.: Extending a fixed-complexity sphere decoder to obtain likelihood information for turbo-MIMO systems. IEEE Trans. Veh. Technol. **57**(5), 2804–2814 (2008)
10. IT++ project. http://itpp.sourceforge.net/current/index.html

Correction to: Energy-Balanced Routing Protocol with QoS Constraints in Ad Hoc Network

Lu Guo, Peng Li, Jiyu Jin, and Jun Mou

Correction to:
Chapter "Energy-Balanced Routing Protocol with QoS Constraints in Ad Hoc Network" in: Q. Liang et al. (Eds.): Communications, Signal Processing, and Systems 2018, Lecture Notes in Electrical Engineering 515, 2019,
https://doi.org/10.1007/978-981-13-6264-4_117

In the original version of the book, the following belated correction has been incorporated: The author name "Filippo, M.D." has been changed to "Di Filippo, M." in reference (12).

The updated version of this chapter can be found at
https://doi.org/10.1007/978-981-13-6264-4_117

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, p. C1, 2019.
https://doi.org/10.1007/978-981-13-6264-4_132

Author Index

Q. Liang et al. (Eds.): CSPS 2018, LNEE 515, pp. 1121–1125, 2019.
https://doi.org/10.1007/978-981-13-6264-4

Z