Lecture Notes in Electrical Engineering

Volume 131

For further volumes:
http://www.springer.com/series/7818

Nabendu Chaki · Natarajan Meghanathan
Dhinaharan Nagamalai
Editors

Computer Networks & Communications (NetCom)

Proceedings of the Fourth International Conference on Networks & Communications

Springer

Editors

Nabendu Chaki
Department of Computer Science
and Engineering
University of Calcutta
Calcutta
India

Dhinaharan Nagamalai
Wireilla Net Solutions PTY Ltd
Albion, VIC
Australia

Natarajan Meghanathan
Department of Computer Science
Jackson State University
Jackson
USA

ISSN 1876-1100 ISSN 1876-1119 (electronic)
ISBN 978-1-4614-6153-1 ISBN 978-1-4614-6154-8 (eBook)
DOI 10.1007/978-1-4614-6154-8
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012953117

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The Fourth International Conference on Networks and Communications (NETCOM 2012) has been held in Chennai, India during December 22–24, 2012. The conference is proved to stimulate researchers from different parts of the world to exchange their ideas in the field of computer networks and data communications including various applications of these. The goal of this conference is to bring together researchers and practitioners from academia and industry to focus on understanding the domain of computer networking and communication technology toward establishing new collaborations in these areas. Authors invited to contribute by submitting original research articles that illustrate research results, projects, survey works, and industrial experiences describing significant advances in the relevant areas.

The conference organizing committee of the NETCOM 2012 took great initiative and interest in circulating the Call for Papers (CFP) for the conference. This effort resulted in a large number of submissions from the researchers of the leading International Universities and Institutes of 32 countries across the world. All the submissions underwent a tough and careful peer-review process with voluntary participation of the committee members and external expert reviewers. The papers had been reviewed based on the novelty of the contributions, technical content, organization, and clarity in presentation. The entire process of initial paper submission, review, and acceptance processes was done electronically. All these efforts undertaken by the Organizing and Technical Committees led to an exciting, rich, and a high-quality technical conference program. The NETCOM 2012 conference featured high-impact presentations for all attendees to enjoy, appreciate, and expand their expertize in the latest developments in Networks and Communications research.

The Technical Program Committee for the conference has selected only 84 papers for publication out of a total 469 number of submissions. The conference proceeding has been organized as a collection of papers presented in the order in which the papers appear in the final technical program for NETCOM 2012. We would like to take this opportunity to thank the General Chairs of NETCOM 2012. We are thankful to all the members of the Technical Program Committee and the external reviewers for their excellent and tireless work. We would also thank Springer for the strong support and the authors who contributed to the success of the conference. Last, but not the least, on behalf of Steering Committee of NETCOM 2012, we sincerely wish that all attendees had been benefited scientifically from the conference. Indeed, we would consider ourselves worth of organizing such an event to offer a platform for all of you if some of tomorrow's works of excellence find its roots initiated from the podium of NETCOM 2012.

Nabendu Chaki
Natarajan Meghanathan
Dhinaharan Nagamalai

Organization Committee

General Chairs

Sanguthevar Rajasekaran	University of Connecticut, USA
Henrique Joao Lopes Domingos	University of Lisbon, Portugal

Steering Committee

Brajesh Kumar Kaushik	Indian Institute of Technology Roorkee, India
Natarajan Meghanathan	Jackson State University, USA
Nabendu Chaki	University of Calcutta, India
Rakhesh Singh Kshetrimayum	Indian Institute of Technology Guwahati, India
Dhinaharan Nagamalai	Wireilla Net Solutions PTY Ltd, Australia
Salah M. Saleh AL-MAJEED	University of Essex, United Kingdom

Program Committee Members

Michal Wozniak	Wroclaw University of Technology, Poland
Jacques DEMERJIAN	Communications and Systems, France
Jan Zizka	SoNet/DI, FBE, Mendel University in Brno, Czech Republic
Sarmistha Neogy	Jadavpur University, India
Yannick Le Moullec	Aalborg University, Denmark
Hwangjun Song	Pohang University of Science and Technology, South Korea
Krzysztof Walkowiak	Wroclaw University of Technology
Ioannis Karamitsos	University of Aegean, Greece
Ramayah	Universiti Sains Malaysia, Malaysia
Khoa N. Le	Griffith School of Engineering, Australia
T. G. Basavaraju	National Institute of Technology Karnataka (NITK), India

Solange Rito Lima	University of Minho, Portugal
Sherif S. Rashad	Morehead State University, USA
Dhinaharan Nagamalai	Wireilla Net Solutions PTY Ltd, Australia
David C. Wyld	Southeastern Louisiana University, USA
Selma Boumerdassi	Cnam/cedric, France
H. V. Ramakrishnan	Bharath University, India
Sattar B. Sadkhan	University of Babylon, Iraq
Eric Renault	Institut Telecom-Telecom SudParis, Evry, France
Alvin Lim	Auburn University, USA
Debasis Giri	Haldia Institute of Technology, India
S. Li	Swansea University, UK
Rushed Kanawati	LIPN-University Paris 13, France
Cristina Ribeiro	University of Waterloo, Canada
Alexander Ferworn	Ryerson University, Canada
Samiran Chattopadhyay	Jadavpur University, India
Rajarshi Roy	IIT, Kharagpur, India
S. A. V. Satyamurty	Indira Gandhi Centre for Atomic Research, India
Laiali Almazaydeh	University of Bridgeport, USA
Shrikant K. Bodhe	Bosh Technologies, India
Alireza Mahini	Islamic Azad University-Gorgan, Iran
N. K. Choudhari	Smt. Bhagwati Chaturvedi College of Engineering, India
Christos Politis	Kingston University, UK
Ayman Khalil	Institute of Electronics and Telecommunications of Rennes, France
Sridharan	CEG Campus-Anna University, India
Mohamed Hassan	American University of Sharjah, UAE
Zuqing Zhu	Cisco Systems, USA
Quan (Alex) Yuan	University of Wisconsin-Stevens Point, USA
Henrique J. A. Holanda	UERN-Universidade do Estado do Rio Grande do Norte, Brazil
Ajay K. Sharma	Dr. B R Ambedkar National Institute of Technology-India
Shrirang A. Kulkarni	National Institute of Engineering, India
Shin-ichi Kuribayashi	Seikei University, Japan
Abdel Salhi	University of Essex, United Kingdom
Antonio Liotta	Eindhoven University of Technology, The Netherlands
Emmanuel Jammeh	University of Plymouth, United Kingdom
Ghaida Al-Suhail	Basrah Univirsity, Iraq
Hao-En Chueh	Yuanpei University, Taiwan, R.O.C
John Woods	University of Essex, United Kingdom
J. K. Mandal	University of Kalyani, India
Ken Guild	University of Essex, United Kingdom
Martin Fleury	University of Essex, United Kingdom

Mohammad M. Banat	Jordan University of Science and Technology, Jordan
Nadia Qadri	University of Essex, United Kingdom
S. Hariharan	J.J.College of Engineering, India
Yasir Qadri	University of Essex, United Kingdom
Wichian Sittiprapaporn	Mahasarakham University, Thailand
Mahi Lohi	University of Westminster, UK
Houcine Hassan	Univeridad Politecnica de Valencia, Spain
Mohammed Ghanbari	University of Essex, United Kingdom
Abdulrahman Yarali	Murray State University, USA
Asmaa Shaker Ashoor	Babylon University, Iraq
Chin-Chih Chang	Chung Hua University, Taiwan
Doina Bein	The Pennsylvania State University, USA
Hossein Jadidoleslamy	University of Zabol, Iran
Kayhan Erciyes	Izmir University, Turkey
M. Mohamed Ashik	Salalah College of Technology, Oman
Mohamed Fahad AlAjmi	King Saud University, Saudi Arabia
Moses Ekpenyong	University of Edinburgh, Nigeria
Natarajan Meghanathan	Jackson State University, USA
Nazmus Saquib	University of Manitoba, Canada
Ruchi Tuli	Yanbu University College, Kingdom of Saudi Arabia
Selwyn Piramuthu	University of Florida, USA
Serguei A. Mokhov	Concordia University, Canada
Rituparna Chaki	West Bengal University of Technology, India
V. Sundarapandian	Vel Tech Dr. RR & Dr. SR Technical University, India
Pinaki Sarkar	Jadavpur University, India
S. Taruna	Banasthali University, India
S. Rajaram	Thiagarajar College of Engineering, India
Uday nuli	Textile and Engineering Institute Ichalkaranji, India
Shun HATTORI	Muroran Institute of Technology, Japan
Yoram Haddad	Jerusalem College of Technology/Ben Gurion University, Israel
Cathryn Peoples	University of Ulster, United Kingdom
Antonio Ruiz-Martinez	University of Murcia, Spain
Paulo R. L. Gondim	University of Brasilia, Brazil
Josip Lorincz	University of Split, Croatia
Jose-Fernan Martinez-Ortega	Universidad Politecnica de Madrid-UPM, Spain
Noor Zaman	King Faisal University, Saudi Arabia
Hangwei	Western Reserve University, USA
Nuno M. Garcia	Universidade Lusofona de Humanidades e Tecnologias, Portugal
Rachida Dssouli	Concordia University, Canada

Jaime Lloret	Polytechnic University of Valencia, Spain
Daqiang Zhang	Nanjing Normal University, China
Juan Jose Martinez Castillo	Ayacucho University, Venezuela
Malamati Louta	University of Western Macedonia, Greece
Malka N. Halgamuge	The University of Melbourne, Australia
Jose Neuman de Souza	Federal University of Ceara, Brazil
Iwan Adhicandra	University of Pisa, Italy
Bob Natale	MITRE, USA
Hamza Aldabbas	De Montfort University, UK
Behnam Dezfouli	University Technology Malaysia (UTM), Malaysia
Ehsan Heidari	Islamic Azad University Doroud Branch, Iran
Jadidoleslamy	University of Zabol, Iran
M. Nadeem Baig	King Saud University, K.S.A
Nisar Hundewale	University of Maryland University College, USA
Omar Almomani	Jadara University, Jordan
Paulo Martins Maciel	Federal University of Pernambuco, Brazil
Phan Cong Vinh	NTT University, Vietnam
Raed alsaqour	Universiti Kebangsaan Malaysia, Malaysia
Sajid Hussain	Fisk University, USA
Sherimon P. C.	Arab Open University, Sultanate of Oman
Somayeh Mohammadi	Islamic Azad University, Iran

Committee Members/Reviewers

A. V. N. Krishna	PJMS CET, India
Vijaya Raju M.	Epoka University (RINAS Campus), Europe
Abbas Aahmad	Hi-Tech College of Engineering and Technology, India
Amandeep Verma	Punjabi University, India
Amanpreet Kaur	ITM University, India
Amitava Mukherjee	BM GBS, India
Anand Kumar	Babasaheb Bhimrao Ambedkar (A Central) University, India
Anitha Vaddinuri	Sree Vidyanikethan Engineering College, India
Anjan K.	RVCE, India
Ankit Agarwal	PICT, India
Ankit Thakkar	Nirma University, India
Annappa	NITK, India
Arvind Kumar Sharma	Sine International Institute of Technology, India
Ashutosh Kumar Dubey	Trinity Institute of Technology and Research, India
B. K. Pattanayak	SOA Deemed to be University, India
Bhaskar Biswas	Banaras Hindu University, India

C. Mala	National Institute of Technology, India
D. Shravani	MIPGS, India
D. Srinivasa Rao	VNRVJIET, India
D. C. Dhubkarya	BIET JHANSI, India
Debabrata Singh	ITER, SOA University, India
Demian Antony D'Mello	St. Joseph Engineering College, India
Dhanalaxmi R.	Anna University Chennai, India
Divya T. V.	MG University India
Ferdinant T.	Jayaram College of Engineering and Technology, India
G. Sankara Malliga	VELS University, India
Gaikwad Dhananjay S.	HSBPVT's Parikrama College of Engineering, India
Gopalakrishnan Kaliaperumal	Anna University, India
Gosta Biswas	Indian School of Mines, India
Himanshu Sharma	GLNAIT Mathura, India
Indumathi	Anna University, India
Jayadev Gyani	Jayamukhi Institute of Technological Sciences, India
Jitendra Maan	Tata Consultancy Services, India
K. Vimala Devi	Kalasalingam University, India
Kahkashan Tabassum	Muffakham Jah College of Engineering and Technology, India
Kamlesh Dutta	National Institute of Technology, India
Kavuri.Roshan	J.B. Institute of Engineering and Technology, India
Khalid NASR	IRIT, India
Koteswara Rao G.	MSBI-HCL, India
Kousik Mukherjee	B.B.College, India
Latha Gannarapu	Kakatiya University, India
M. Upendra Kumar	Mahatma Gandhi Institute of Technology (MGIT), India
Mala	National Institute of Technology, India
Manjula Shenoy K.	MIT, India
Md. Mahmudul Hasan	Daffodil International University, Bangladesh
Minal moharir	R V College of Engineering, India
Mohd Dilshad Ansari	Invertis University, India
Mrinal Naskar	Jadavpur University Kolkata, India
Muttanna Kadal H. K.	Dr. AIT, India
N. Bhalaji	PERI Institute of Technology, India
Naishita Taraka	JNTU, India
Nandini Mukherjee	Jadavpur University, India
Neetesh	Indian Institute of Technology Indore, India
Nishant doshi	National Institute of Technology, India
Nityananda Sarma	Tezpur University, India

PESN. Krishna Prasad	Prasad V. Potluri Siddhartha Institute of Technology, India
P. Perumal	Sri Ramakrishna Engineering College, India
P. R. S. M. Lakshmi	Vignan University, India
Pankaj Sharma	ABES Engineering College Ghaziabad, India
Parveen kumar	Lovely Professional University, India
Poonam Garg	Institute of Management Technology, India
Pradeep	Sri Venkateshwara College of Engineering and Technology, India
Pranay Meshram	St. Vincent Palloti College of Engineering and Technology, India
Prasad Halgaonkar	MIT College of Engineering, India
Preetee K. Karmore	YCCE-Engineering College, India
R. Venkadeshan	Chettinad College of Engineering and Technology, India
R. Deepa	Amrita Vishwa Vidyapeetham, India
R. Selvarani	M S Ramaiah Institute of Technology, India
Racing Ruso	Anna University, India
Rahul Johari	Guru Gobind Singh Indraprastha University, India
Rajeshwari Hegde	BMS College of Engineering, India
RavendraSingh	MJP Rohilkhand University, India
Revathi	SRM University, India
Richard William	Jayalakshmi Institute of Technology, India
Ripal Patel	BVM Engineering College, India
S. Britto	Bharathidasan University, India
S. K. V. Jayakumar	Pondicherry University, India
Sandeep M. Chaware	D.J. Sanghvi College of Engineering, India
Sandhya Magesh	B.S.Abdur Rahman University, India
Santhi Thilagam	NITK Surathkal, India
Shahid Siddiqui	Integral University, India
Shaik Sahil Babu	Adavpur University, India
Sharmila Sankar	BSA Abdur Rahman University, India
Shatheesh sam	Nesamony Memorial Christian College, India
Shivaputra	Dr. Ambedkar Institute of Technology, India
Shriram K. Vasudevan	Amrita University, India
Siddesh G. M.	M.S. Ramaiah Institute of Technology, India
Soubhik Chakraborty	Birla Institute of Technology, India
Soumen Kanrar	Vehere Interactive Pvt Ltd, India
Sowmya	NITK, India
Srinivas	Jyothishmathi institute of Technology and Science, India
Subhrendu Guha Neogi	Sir Padampat Singhania University, India
Sunil Kumar Gupta	BCET GURDASPUR, India
T. Meyyappan	Alagappa University, India

T. P. Surekha	Vidya Vardhaka College of Engineering, India
Tapalina Bhattasali	West Bengal University of Technology, India
Tumpa Roy	GLA Groups of Institutions, India
Urmila Shrawankar	G H Raisoni College of Engineering, India
Utpal Biswas	University of Kalyani, India
V. Jayalakshmi	Sudharsan Engineering College, India
Vasu K.	IIT Kharagpur, India
Vijay H. Mankar	Government Polytechnic, India
Vikas J. Dongre.	Government Polytechnic, India
Vivekanand Mishra	S.V. National Institute of Technology, India
Vivekanandan Mahadevan	SRM University, India
Y. Srinivasa Rao	Andhra University College of Engineering, India
Y. Venkataramani	Saranathan College of Engineering, India
Zeenat Rehena	Jadavpur University,India
Zulfa Shaikh	Acropolis Institute of Technology and Research, India
Srinivasarao	Defence University College, Ethiopia
Zheng Chang	University of Jyvaskyla, Finland
Mohammad Zunnun Khan	Integral University, India
Gaurav Somani	LNMIIT, India
Mohammad	University of Botswana, Botswana
M. Sandhya	B.S. Abdur Rahman University, India
Elboukhari Mohamed	University Mohamed First, Morocco
Durgesh Samadhiya	Chung Hua University, Taiwan
Abbas Ahmad	Hi-Tech College of Engineering & Technology, India
Anu Bala	Chandigarh Engineering College, India
B. Jagadeesh	G.V.P. College of Engineering, India
Er. Saba Khalid	Integral University, India
Gagan Jindal	Chandigarh Engineering College, India
Hameem Shanavas	MVJ College of Engineering, India
Indrajit Banerjee	BESU, Shibpur, India
K. Kishan Rao	Vaagdevi Group of Technical Institutions, India
K. Suganthi	Madras Institute of Technology, India
Kaushik	IIT Kharagpur, India
Neeraj Kumar	Thapar University, Patiala (Punjab), India
P. Suresh Varma	Adikavi Nannaya University, India
Prabu D.	NetApp Inc., India
Prasun Chowdhury	Jadavpur University, India
R. R. Mudholkar	Shivaji University, India
Rajashree Biradar	Bellary Institute of Technology, India
RaviShankar Yadav	CAIR DRDO, India
Salil Kumar Sanyal	Jadavpur University, India
Samarendra Nath Sur	Sikkim Manipal Institute of Technology, India
Sanjay M. Koli	E & TC Department SKNCOE, India

Sarbani Roy	Jadavpur University, India
Seema Verma	Banasthali University, India
Shailaja kanawade	Sandip Institute of Technology and Research Centre, India
Sudesh	Sri Mata Vaishno Devi University, India
Vanathi B.	Anna University Chennai, India

Contents

Part I The Fourth International Conference on Networks & Communications (NETCOM-2012): Adhoc and Sensor Networks

Part III The Fourth International Conference on Networks & Communications (NETCOM-2012): Measurement and Performance Analysis

Part IV The Fourth International Conference on Networks & Communications (NETCOM-2012): Network Architectures, Protocols and Routin0067

Part V The Fourth International Conference on Networks & Communications (NETCOM-2012): Network Operations and Management

Part VI The Fourth International Conference on Networks & Communications (NETCOM- 2012): Network Security, Trust and Privacy

Part VII Workshops: The Fourth International Workshop on Network and Communications Security (NCS 2012)

Part VIII Workshops: The Fourth International Workshop on Wireless and Mobile Networks (WiMoNe-2012)

Part IX Workshops: Fourth Workshop on Applications of Graph Theory in Wireless Ad hoc Networks and Sensor Networks (GRAPH-HOC - 2012)

Part X Workshops: The Fourth International Workshop on Ad Hoc and Ubiquitous Computing (AUC- 2012)

Part I
The Fourth International Conference on Networks & Communications (NETCOM-2012): Adhoc and Sensor Networks

Perspectives of Sybil Attack in Routing Protocols of Mobile Ad Hoc Network

Manu Sood and Amol Vasudeva

Abstract It is cumbersome to achieve the security in a mobile ad hoc network due to its open nature, dynamically changing topology, lack of infrastructure and central management. A particular harmful attack that takes the advantage of these characteristics is the Sybil attack, in which a malicious node illegitimately claims multiple identities. Two routing mechanisms vulnerable to the Sybil attack in the mobile ad hoc networks are multi-path routing and geographic routing. In addition to these routing protocols, we show in this paper that the Sybil attack can also disrupt the head selection mechanism of various cluster-based routing protocols such as lowest id, highest node degree and mobility based clustering. To achieve this, we illustrate to have introduced a category of Sybil attack in which the malicious node varies its transmission power to create a number of virtual illegitimate nodes called Sybil Nodes, for the purpose of communication with legitimate nodes of the MANETs.

Keywords Mobile ad hoc network · Sybil attack · Malicious node · Sybil node · Network security · Routing protocol

1 Introduction

Security is an important concern in the Mobile Ad hoc Networks (MANETs). However, the characteristics of MANETs pose both challenges and opportunities in achieving the security goals, such as confidentiality, authentication, integrity,

M. Sood
Department of Computer Science, Himachal Pradesh University,
Shimla, Himachal Pradesh, India

A. Vasudeva (✉)
Department of Computer Science and Engineering and Information Technology,
Jaypee University of Information Technology, Waknaghat,
Solan, Himachal Pradesh, India
e-mail: amol_dev@rediffmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_1,

availability, access control and non-repudiation [1]. There are a wide variety of attacks that target the weakness of MANET routing protocols. Most sophisticated and subtle routing attacks have been identified in some recently published papers such as Blackhole [2], Rushing [3], wormhole [4] and Sybil attack [5] etc. A Sybil attack is an attack [5], in which a malicious node illegally claims multiple identities by impersonating other nodes or by claiming fictitious identities. Sybil attacks are also capable of disrupting the routing mechanisms in mobile ad hoc networks. Karlof and Wagner have shown in [6] that multi-path routing and geographical routing schemes are affected by this attack. In case of multi-path routing a set of supposedly disjoint paths can all be passing through the same malicious node, which is using several Sybil identities. Also in location based routing a malicious node can present multiple Sybil nodes with different positions to its neighbors. Therefore, a legitimate node may choose any of the Sybil nodes while forwarding the packet on the basis of nearest location to the destination node; but in reality it will be passing the packets through the malicious node.

In addition to these routing protocols, we have shown in this paper that the Sybil attack can also disrupt the head selection mechanism of various cluster-based routing protocols such as lowest ID [7], highest node degree [8] and mobility based clustering [9]. To the best of our knowledge, this is for the first time that the impact of Sybil attack has been shown in these cluster based routing algorithms. The rest of this paper is organized as follows. Section 2 describes the Sybil attack in details. Section 3 describes the following routing protocols: Split Multi-path Routing (SMR) [10], Greedy Perimeter Stateless Routing (GPSR) [11] and various cluster based routing protocols along with the effect of Sybil attack on these protocols, respectively. Finally, the Sect. 4 concludes the paper.

2 Sybil Attack

Sybil attack was first introduced by J. R. Douceur. According to Douceur, the Sybil attack is an attack by which a single entity can control a substantial fraction of the system by presenting multiple identities [5]. The Sybil attack can occur in a distributed system that operates without a central authority to verity the identities of each communicating entity [12].

In a Mobile Ad hoc Network, the only way for an entity to detect the presence of other entities is by sending and receiving the messages over a shared broadcast communication channel. By taking the advantage of this feature, a malicious node can send messages with multiple fake identities. The node spoofing the identities of the nodes is called malicious node/Sybil attacker, and the nodes whose identities are spoofed are called Sybil nodes. Figure 1 represents a malicious node S along with its four Sybil nodes (S_1, S_2, S_3 and S_4). If this malicious node communicates with any legitimate node by presenting all its identities, the legitimate node will have illusion that it has communicated with five different nodes. But in actual, there exists only one physical node with multiple different IDs. If a single malicious node is able to

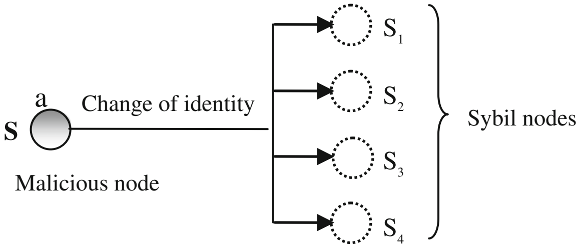


Fig. 1 A Sybil attacker with multiple IDs

convince its neighbors by presenting multiple identities, it will have control over the substantial portion of the network and can adversely affect the functioning of the network. According to Newsome [12], the mechanisms that are affected by the Sybil attack are: Data Aggregation, Fair Resource Allocation, Voting and Misbehavior Detection etc. Karlof and Wagner have shown in [6] that Sybil attack can also disrupt the functioning of certain routing protocols in MANETs such as multi-path routing protocols [10, 13, 14] and geographic based routing protocols [11, 15, 16].

The launching of the Sybil attack can be represented using three dimensions: Communication, Participation and Identity [12]. Newsome et al. state that there are two ways of communication: Direct and Indirect [12]. In a direct communication, as the name implies, the malicious node allows its Sybil nodes to communicate directly with the legitimate nodes. In case of indirect communication, the malicious node does not allow its Sybil nodes to communicate directly with the legitimate nodes. However, the authors are of the opinion that the title 'Establishment of the Connection' would have been more appropriate instead of the title 'Communication'. Participation is concerned about the participation of Sybil nodes in the communication with legitimate nodes in the network. These nodes can participate simultaneously or non-simultaneously. There are two methods by which a Sybil node can get the identity: In the first method a Sybil node can steal the identity of a legitimate node by impersonating it. The second method involves the fabrication of a fresh fake identity.

3 Sybil Attack in MANET Routing Protocols

In this section, we have illustrated the impact of Sybil attack on Split Multi-path Routing (SMR) and Greedy Perimeter Stateless Routing (GPSR). In addition to these routing protocols, we have shown that the Sybil attack can also disrupt the different forms of Cluster Based Routing Protocols such as Lowest ID Clustering, Highest Node Degree Clustering and Mobility based Clustering.

3.1 Sybil Attack in Split Multi-Path Routing (SMR)

Split Multi-path Routing (SMR) [10], one of the multi-path routing protocols based on Dynamic Source Routing (DSR) [17], establishes and utilizes multiple maximally disjoint paths. Unlike DSR, the intermediate nodes in SMR do not respond to route requests, in order to obtain maximal node disjoint paths. Intermediate nodes forward the first RREQ they receive and instead of dropping all the duplicate RREQ packets, rebroadcast those duplicate packets that are being received through a different incoming link and whose hop count is not greater than the previously received RREQs. When the destination node receives the first RREQ, it responds with RREP to the source node and then waits for certain duration of time, to receive additional requests. The destination node then selects the route that is maximally disjoint to the route that is already replied. Consider Fig. 2a, where the source node S floods the RREQ packets to find an optimal route to the destination node D. The intermediate nodes forward the duplicate RREQ packets that traversed through a different incoming link than the link from which the first RREQ is received, and whose hop count is not greater than that of the first received RREQ. Now assume that a Sybil attacker node M has established itself in the network with two fake IDs i.e. X and Y. Thus, in this case the packets are being forwarded through a single physical node i.e. M. When

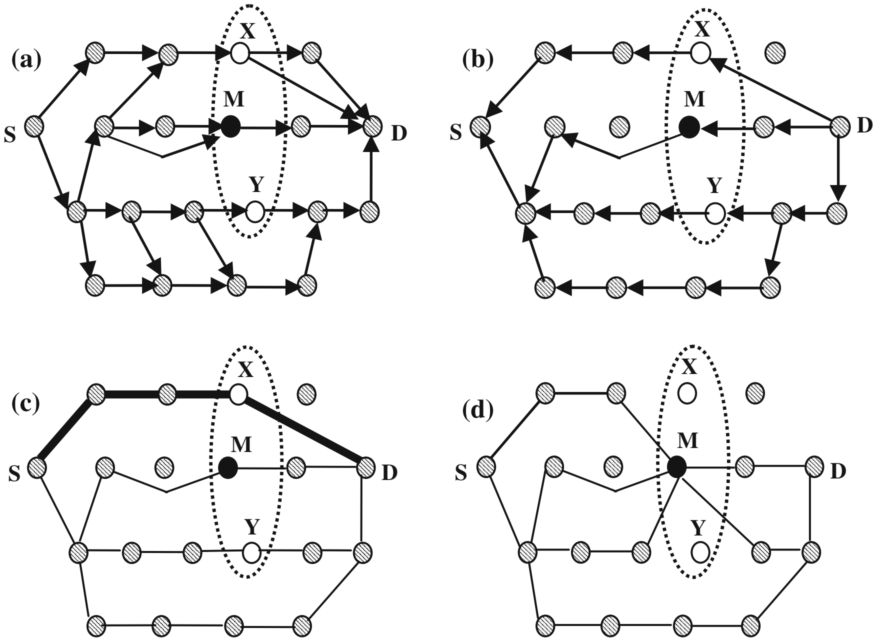


Fig. 2 **a** Flooding of RREQ packet from S to D. **b** RREP packets from D to S. **c** Selections of multiple disjoint paths. **d** Routes passing through the same node M. **S**-Source Node, **D**-Destination Node, **M**-Malicious Node, **X**, **Y**-Sybil Nodes

the RREQ packets have started to arrive at the destination node, it starts to send the RREP packets back towards the source node (Fig. 2b).
The path followed by each RREP packets is same as that of their corresponding RREQ packets. After receiving all the RREP packets the source node S makes the entry in its routing table. It then chooses the route with minimum hop count of 4 as shown in Fig. 2c by a thick line. But in reality, three routes are being passed through the same malicious node M and the Fig. 2d depicts that the routing mechanism has been disrupted, effectively.

3.2 Sybil Attack in GPSR Routing

The Greedy Perimeter Stateless Routing (GPSR) algorithm [11] works in two modes: greedy forwarding and perimeter forwarding. The algorithm starts forwarding with the greedy mode, by default. In greedy forwarding the source node looks for its neighbor that is closest to the destination and forwards the packet to that node. This process is repeated for the next node also and so on. Consider a MANET topology with 11 nodes as shown in the Fig. 3. Assume that the node C is a malicious node or Sybil attacker who somehow has succeeded to enter into the network. The actual position of this node is (7, 9). This node has presented two fake IDs i.e. the Sybil nodes D, E with their locations as (4, 11) and (11, 8), respectively. The node communicates with the other neighbors in its region by providing all the three locations (one actual and two fake). Thus, an adversary may claim to be present at more than one location for its neighbors by sending multiple HELLO messages, each time with different location information. Now suppose that the node A wants to send the packet to the destination I and to find the route it follows the greedy forwarding. A’s radio range is denoted by the dotted circle about A and the arc with the radius equal to the distance between A and I is shown as arc about I. The node A forwards the packet to the Sybil node E, as it finds that the distance between E and I is less than between E and any of the A’s other neighbors, according to the location information available in its Table 1. But in fact, node A has forwarded the packet to node C whose location is (7, 9) having a distance greater than the distance from node D (8, 2). Therefore, the routing scheme has been disrupted in the MANET with the entry of Sybil attacker.

3.3 Sybil Attack in Cluster-Based Routing Protocols

A Sybil attacker can also send the messages by varying its transmission power for all of its identities. The advantage of varying the transmission power for all the Sybil nodes is that the received signal strength at the receiving node with respect to these Sybil nodes will also be different. We have used the feature of variable transmission power to show that the Sybil attack can also disrupt various cluster based routing schemes such as lowest ID, highest node degree and mobility based clustering.

Nodes with their current locations

A	(6,7)
B	(8,2)
C	(7,9)
D	(4,11)
E	(11,8)
F	(13,11)
G	(14,16)
H	(15,10)
I	(20, 10)
J	(17,8)
K	(15,6)

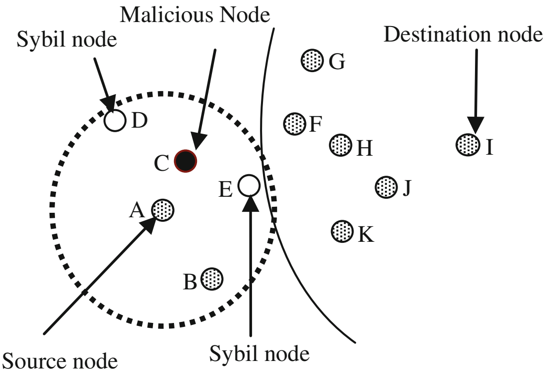


Fig. 3 A MANET topology with 11 nodes

Table 1 Location information of A's one hop neighbors

Neighbors of the source node A	Locations of A's neighbors	Distance of A's neighbors from the destination node I [10, 20]
B	(08, 02)	14.42
C	(07, 09)	13.04
D	(04, 11)	16.03
E	(11, 08)	09.22

Sybil Attack in Lowest ID Clustering Algorithm. In lowest ID algorithm [7], a node with the minimum ID is chosen as a clusterhead. Each node is provided with a unique ID and it periodically broadcasts the list of its neighbor's IDs, including itself. A node which only hears nodes with ID higher than itself is a clusterhead (CH).

Figure 4a shows a schematic of the result of using lowest ID clustering. There are 11 nodes with unique IDs, which form a connected graph. After the Lowest-ID clustering algorithm is executed, three clusters are formed, as depicted by the dotted circles. The black colored balls inside each cluster represent the clusterheads (1, 5 and 3 in Fig. 4). The striped balls (6 and 7) that are within the communication range of two or more different clusters represent the gateway nodes and the empty balls are the member nodes.

To become a cluster head, a malicious node can present the Sybil node with lowest ID in its neighborhood. For this, the malicious node will have to behave normally for the period until it has accessed the information about the whole network i.e. its one-hop, two-hop and n-hop neighbors and their respective IDs. After gaining the appropriate information, the malicious node can introduce its Sybil node with lowest ID, to fulfill its purpose by becoming the clusterhead. The attack becomes more devastating and difficult to be detected if the malicious node introduces its fake identities (Sybil nodes) by varying the transmission power. It takes the advantage of varying the transmission power in two ways: First, it cannot be detected on the basis

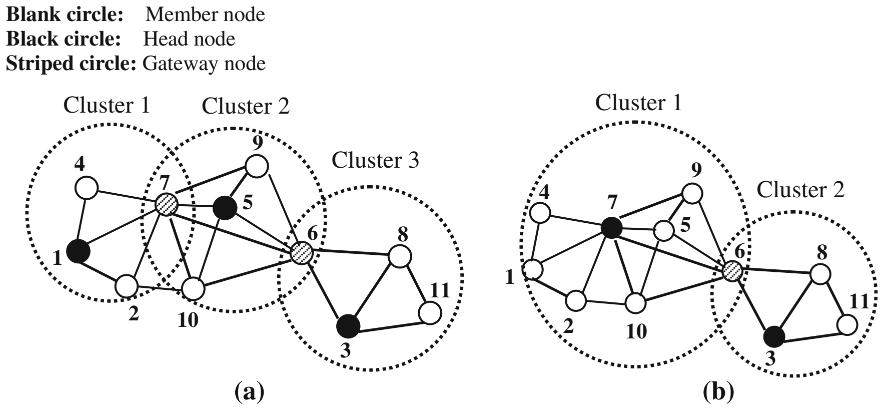


Fig. 4 Cluster formation process. **a** Lowest ID approach. **b** Highest node degree approach

of same signal strengths of its Sybil nodes [18]. Second, by decreasing the transmission power for different Sybil nodes, the message will not reach all the legitimate neighbors of the malicious node and hence cannot be detected on the basis of the fact that two different physical entities cannot have the same set of neighbors [19].

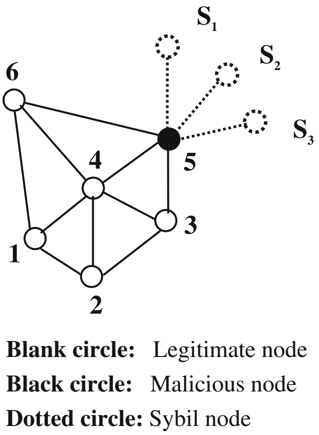
The Sybil attack can also disrupt the lowest ID based cluster routing by presenting multiple Sybil nodes with IDs higher than its neighboring legitimate nodes. Here the intention is to make the legitimate node with lowest ID, the clusterhead again and again to drain its battery. Once the battery is drained completely, the malicious node can impersonate its ID for one of its Sybil node to become a clusterhead.

Sybil Attack in Highest Node Degree Algorithm. In highest node degree algorithm [8], the degree of a node is computed on the basis of its neighbors. The node having maximum number of neighbors is elected as the clusterhead. If there is a tie between two or more nodes in terms of node degree, the node with lowest ID is chosen to be clusterhead. Figure 4b shows the result of using highest degree clustering for the same topology that was being used for the lowest ID algorithm.

The highest degree algorithm can also be disrupted by the Sybil attack. By presenting multiple Sybil nodes, a malicious node may claim to have more neighbors than the actual number. For example, in the Fig. 5, the node 5 is a malicious node with nodes 3, 4 and 6 as its one hop neighbors (hence node degree 3). The node 4 has the maximum node degree of 5 among its neighbors and should to be selected as a cluster head. But, the malicious node 5 also includes its three Sybil node, i.e. S_1, S_2 and S_3 so as to increase its node degree to 6 and hence becomes the clusterhead. Now the question is how to introduce these Sybil nodes to the legitimate neighboring nodes, i.e. with direct communication or indirect communication.

If indirect method of communication is followed, the malicious node will claim to have the specified number of Sybil identities as its neighbors and will not allow them to communicate directly with its legitimate neighboring nodes. But, due to mobile

Fig. 5 A MANET topology with 6 nodes


S1
S2
S3
6
4
5
3
1
2
Blank circle: Legitimate node
Black circle: Malicious node
Dotted circle: Sybil node

nature of the MANET, it is not possible for these Sybil nodes to be in the transmission range of only a single node for the whole life of this MANET. If this trick is used by the malicious node repeatedly, the Sybil attack can be detected on suspicion of not communicating directly with the other legitimate nodes in the network for a long time. Also, the Sybil attack can be detected on the basis of movement of same set of nodes, by the observer nodes [20].

Therefore, to win the trust of legitimate neighboring nodes, a malicious node should allow its Sybil nodes to communicate directly with these legitimate neighboring nodes. But the problem in a direct communication is that the node degree of legitimate neighboring nodes will also increase by the same factor as that of the malicious node. For example, in Fig. 5, if a malicious node 5 allows its 3 Sybil nodes to communicate directly with its neighboring nodes, its own node degree becomes $(3+3) = 6$. But this will also lead to increase in the node degrees of its neighboring nodes 3, 4 and 6 by the same factor, i.e. 3. As a result the nodes degree of the node 4 becomes $(5 + 3) = 8$, and should be selected as the cluster head according to the algorithm.

A better option is to allow all the Sybil nodes to communicate directly with the neighboring nodes, by decreasing their transmission powers. After decreasing the transmission power, hello packets sent by the Sybil node will reach only to a subset of neighboring nodes that are closer to the malicious node. As a result, there will be increase in the node degree of this particular subset of neighbors of the malicious node, only. Before the inclusion of Sybil nodes, if the node degree of each node in this subset was less than the node degree of malicious node, the malicious node will become a clusterhead. Otherwise, one of the legitimate nodes present in that subset may become a clusterhead. Therefore, the probability of a malicious node to become a clusterhead is less in this scheme.

In order to increase the chances of becoming a clusterhead, the malicious node may claim to have some additional Sybil nodes by communicating them, indirectly. Therefore, in this scheme the Sybil nodes are kept in two different pools: one pool of

Sybil nodes for direct communication and the other pool of Sybil nodes for indirect communication. The malicious node will also keep on exchanging the Sybil nodes of these pools continuously, which makes the detection of Sybil attack very difficult.

Sybil Attack in Mobility based Clustering. Basu et al. in [9] have proposed an algorithm for the mobility based clustering (MOBIC) approach in MANETs. This algorithm uses mobility of the nodes as a feature to form the clusters. Each node in the Mobile Ad hoc Network computes the ratio of two successive "Hello" messages from all its neighbors. This gives the relative mobility metric of the nodes with respect to each of their respective neighbors. Then, by calculating the variance of relative mobility values of all the nodes with respect to their neighbors, the aggregate speed of all the mobile nodes can be estimated. Finally, the mobile node with lowest variance value in its neighborhood is elected as the clusterhead.

This algorithm can also be affected by the Sybil attack. For example, consider that we have to find the aggregate mobility of a node 4 with respect to its neighboring nodes 1, 2, 3, 5 and 6 as shown in the Fig. 5. Let the node 5 be a malicious node with three Sybil nodes S_1, S_2 and S_3. Therefore, in addition to itself the malicious node will also use its different Sybil nodes to send consecutive Hello messages, by varying its transmission power. The variation in the transmission power will be adjusted in such a manner that the second received signal strength at the legitimate receiving nodes is comparatively less than the first one. After computing the relative mobility values of all the neighboring nodes (legitimate nodes and Sybil nodes), the aggregated mobility of the node 4 is calculated by taking the variance of relative mobility values of all its neighboring nodes. The contribution of the Sybil nodes is only to increase the variance of the node 4, which is the measure of aggregated relative mobility. That is, greater the variance of a node, lesser is the chance for that node to become the clusterhead. The same process will be repeated for other legitimate nodes in the neighbor of the malicious node. Thus, the probability that the malicious node will become a clusterhead, will increase in this manner.

4 Conclusion and Future Work

In this paper, we have highlighted the impact of the Sybil attack on different routing schemes in the MANETs, such as multi-path routing, location based routing and cluster based routing. We have shown that a Sybil attack can disrupt the head selection mechanism of various cluster-based routing protocols, such as lowest ID, highest node degree and mobility based clustering. In the lowest ID clustering algorithm, a malicious node can present a Sybil node with lowest ID in its neighborhood to become the clusterhead. The attack becomes very difficult to detect and more destructive if the malicious node introduces its fake identities (Sybil nodes) by varying the transmission power. In the highest degree clustering algorithm, a malicious node may claim to have more number of neighbors by presenting multiple Sybil nodes. To achieve this, a malicious node keeps its Sybil nodes in two different pools. The Sybil nodes in the first pool are allowed to communicate directly with the neighboring nodes, by decreasing their transmission powers. As a result, the hello packets sent by these Sybil nodes will not be received by all the neighbors of the malicious node. Therefore, there will be increase in the node degree to only a subset of neighbors that are closer to the malicious node. Prior to introduction of the Sybil nodes, if the node degree on any legitimate node in this subset was greater or equal to the node degree of the malicious node, then the chances for this malicious node to become a clusterhead are very low. Therefore, to increase the chances, this malicious node may claim to have some additional Sybil nodes by introducing them, indirectly. To disrupt the mobility based clustering scheme such as MOBIC, a malicious node uses its different Sybil node to send consecutive Hello messages, by varying its transmission power. The variation in the transmission power is adjusted in such a manner that the second received signal strength at the legitimate receiving nodes is comparatively less than the first one. Therefore, the malicious node along with its Sybil nodes will contribute in increasing the variance of its legitimate neighboring nodes, which decreases their probability to become a clusterhead. One of the objectives of this study is to have a better understanding of challenges offered by the Sybil attack on this routing protocol. Presently we are in the process of designing an appropriate Sybil attack detection mechanism. The credibility and efficiency of this mechanism will be tested for the various forms of Sybil attack, using the network simulator.

References

1. Wu B, Chen J, Wu J, Cardei M (2006) A survey on attacks and countermeasures in mobile ad hoc networks. In: Xiao Y, Shen X, Du D-Z (eds) Wireless/mobile network security. Springer, New York, pp 103–135
2. Al-Shurman M, Yoo S-M, Park S (2004) Black hole attack in mobile ad-hoc networks. In: ACM Southeast Regional conference
3. Hu YC, Perrig A, Johnson DB (2003) Rushing attacks and defense in wireless ad hoc networks routing protocol. In: Proceedings of ACM WiSe2003
4. Hu Y, Perrig A, Johnson D (2002) Packet leashes: a defense against wormhole attacks in wireless ad hoc networks. In: Proceeding of IEEE INFORCOM
5. Douceur JR (2002) The Sybil attack. IPTPS '01: revised papers from the first international workshop on peer-to-peer systems. Springer Verlag, London, UK, pp 251–260
6. Karlof C, Wagner D (2003) Secure routing in wireless sensor networks: attacks and countermeasures. Elsevier's Ad Hoc Netw J Special Issue Sens Netw Appl Protoc 1(2–3):293–315
7. Ephremides A, Wieselthier JE, Baker DJ (1987) A design concept for reliable mobile radio networks with frequency hopping signaling. Proc IEEE 75(1):56–73
8. Gerla M, Tsai JTC (1995) Multicluster, mobile, multimedia radio network. ACM/Baltzer Wirel Netw J 1:255–265
9. Basu P, Khan N, Little T (2001) A mobility based metric for clustering in mobile ad hoc networks. In: Proceedings of the 21st international conference on distributed computing systems workshops (ICDCSW '01), pp 413–418
10. Lee S-J, Gerla M (2001) Split multipath routing with maximally disjoint paths in ad hoc networks. In: Proceedings of the IEEE international conference on communications (ICC). Helsinki, Finland, pp 3201–3205, June 2001

11. Karp B, Kung HT (2000) GPSR: greedy perimeter stateless routing for wireless networks. In: Proceedings of IEEE/ACM MobiCom, pp 243–254
12. Newsome J, Shi E, Song D, Perrig A (2004) The Sybil attack in sensor networks: analysis and defenses. In: Proceedings of the 3rd international symposium on information processing in sensor networks (IPSN '04), ACM, Berkeley, California, USA, pp 259–268
13. Lee S-J, Gerla M, Chiang CC (1999) On-demand multicast routing protocol (ODMRP). In: Proceedings of IEEE WCN'99, Sep 1999
14. Royer EM, Perkins CE (1999) Multicast operation of ad hoc on-demand distance vector routing protocol. In: Proceedings of ACM MOBICOM, pp 207–18, August 1999
15. Basagni S, Chlmtac I, Syrotiuk VR, Woodward BA (1998) A distance routing effect algorithm for mobility (DREAM). In: Proceedings of IEEE/ACM MobiCom, pp 76–84
16. Ko Y, Vaidya NH (1998) Location-aided routing (LAR) in mobile ad hoc networks. In: Proceedings of IEEE/ACM MobiCom, pp 66–75
17. Johnson DB, Maltz DA (1996) Dynamic source routing in ad hoc wireless networks. In: Imielinski T, Korth H (eds) Mobile computing. Kulwer Academic Publishers, Boston, pp 153–181
18. Demirbas M, Song YW (2006) An RSSI-based scheme for Sybil attack detection in wireless sensor networks. In: International workshop on wireless mobile multimedia. New York, USA, pp 564–570
19. Ssu K-F, Wang W-T, Chang W-C (2009) Detecting Sybil attacks in wireless sensor networks using neighboring information. Comput Netw 53(18):3042–3056
20. Piro C, Shields C, Levine BN (2006) Detecting the Sybil attack in mobile ad hoc networks. In: Securecomm and workshops

A Jini Based Implementation for Best Leader Node Selection in MANETs

Monideepa Roy, Pushpendu Kar and Nandini Mukherjee

Abstract MANETs provide a good alternative for handling the constraints of disconnectivity and disruption of static communication infrastructure faced during emergency situations. In such frameworks, it sometimes becomes necessary to elect a leader node for the MANET. Electing a leader node in a MANET poses several challenges due to the inherent properties of mobility, resource constraints etc. of these ad hoc networks. Although there are at present some existing leader node selection algorithms, it is difficult to apply them to MANET based frameworks where multiple selection of leader nodes is allowed, because of various reasons. This paper presents an algorithm for the best leader node selection, along with a network wide searching technique that a client uses to search for a particular service in the network, both of which are especially suitable for these types of frameworks.

Keywords MANETs · Leader node selection · Jini

M. Roy (✉)
School of Computer Engineering, KIIT University,
Bhubaneshwar, Odisha, India
e-mail: monideepa.roy@gmail.com

P. Kar
Department of Computer Science and Engineering, Jadavpur University,
Kolkata, West Bengal, India
e-mail: pushpendukar@rediffmail.com

N. Mukherjee
Department of Computer Science and Engineering, Jadavpur University,
Kolkata, West Bengal, India
e-mail: nmukherjee@cse.jdvu.ac.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_2,

1 Introduction

Leader election in a network is the process of electing a node uniquely as the leader of that system. There is a number of existing leader election algorithms which have been proposed for classical distributed systems. But they are not suitable for MANETs which do not require any infrastructure or central control nodes, since they have independent network organization, and dynamic topology. And the participating mobile devices in a MANET also suffer frequently from constraints of resources and loss of communication due to their mobility. So leader election is challenging in MANETs. There are several existing leader election algorithms that have been proposed for ad hoc networks also. But the assumption in most of the cases is that the nodes are homogeneous and only one node is selected as the leader for the network. However, in our proposed framework, there may be multiple leader nodes present simultaneously in the framework. So a service provider or client basically has to choose to register with the most suitable leader node from among the multiple leader nodes within the range. In this paper we have presented an algorithm for a service provider or a client to register with the most suitable leader node based on factor called the threshold value of a leader node, the calculation of which is given in Sect. 4. This paper also presents a network wide search process which is performed by the clients while searching for a particular service. The rest of this paper is organized as follows: Sect. 2 explains the justification for proposing this algorithm, Sect. 3 presents the related work in this area. Section 4 presents the Leader node selection algorithm and Sect. 5 describes the network wide search activity. Finally, Sect. 6 contains the conclusion and future work.

2 The Need for Best Leader Node Selection

The standard definition of the leader election problem for static networks is that: I. Eventually there is a leader and II. There should never be more than one leader. Since in mobile ad hoc networks, the topology may change frequently, so, in such circumstances, it is necessary to reallocate the task of a leader node to another competent node. The existing leader election algorithms for MANETs assume that all the nodes are homogeneous. However, our work focuses on a mobile Grid framework which is composed of heterogeneous nodes (laptops, PDAs and other types of mobile devices) with different capabilities. Furthermore, complications arise because partitions can occur in an ad hoc network due to mobility of nodes. In such a case, some applications require that every component of the partition must have a unique leader. The mobile Grid framework which we had presented in our earlier work, is MANET—based and uses the Jini network technology, where the nodes are classified as being of three types: leader nodes, client nodes and service provider nodes. The nodes that are capable of being designated here as the leader nodes, are actually the nodes which host the Jini lookup service. A lookup service acts as repository of services, where a proxy object of each registered service of a service provider is stored with a set of attributes which define the service. The service provider nodes are mobile devices which have some services to offer and register with a leader node. The registration consists of acquiring a lease [1] from the lookup service and letting the lookup service know what services they have to offer (if any). The duration of the lease is to be granted to a service provider is decided by the lookup service based on various parameters. The clients also register with any one of the leader nodes to avail a service. If a requested service is not provided by the leader node that the client has registered with, then that leader node retrieves details of similar services from nearby leader nodes. This framework allows multiple leader nodes to be accommodated within a network and with overlapping ranges. Therefore if a service provider or client is within the range of multiple leader nodes, the proposed Jini [2] based algorithm is applied to select the most suitable leader node it should register with.

3 Related Work

While there are a number of leader election algorithms for traditional networks, they do not work in the highly dynamic environment found in mobile networks. This is primarily because these solutions to the leader election problem assume a static topology or assume that topological changes stop before an election starts or assume an unrealistic communication model such as a message-order preserving network. Leader election algorithms for mobile ad hoc networks have been proposed also [3–5]. But they have certain shortcomings for which they are not suitable for MANET based frameworks [6]. Firstly, these algorithms are designed to perform random node elections. Secondly, the existing algorithms assume that all the nodes in the network are homogeneous (i.e. having the same configurations) and the election is made from all the participating nodes in the network. Thirdly, most existing algorithms assume at most one leader node in the network. Because of the above factors, unlike existing work on leader election we present a leader node selection algorithm in a mobile wireless setting. In our framework [6], some nodes which have relatively higher resource and computational power are pre-designated as leader nodes when they join the network. Therefore all the participating nodes may not be eligible for the election. This algorithm calculates the minimum threshold value for each leader node. Since our framework supports multiple leader nodes, this algorithm chooses the most suitable node for each service provider. In fact we are dealing with the subsequent scenario of when a service provider or a client has to register with a leader and is faced with multiple options of leader nodes. In this case the selection of the best leader node for a service provider/client is not done from all the leader nodes in the network, but only from among the leader nodes which are within signal range of the service provider/client.

4 The Leader Node Election Algorithm

We consider any predefined operational area. Within this area a number of service providers, clients and leader nodes are deployed. The calculations are based on the assumption that the mobile devices do not leave the boundary of the total operational coverage area but may move around within it for the duration of time the application is running. So at any time *t* the position coordinates of the device can be determined. The pattern of movement of the nodes is assumed to be the Random Waypoint Mobility Model [7]. In this mobility model a node moves from its current location to a new location by randomly choosing a direction and speed in which to travel. The new speed and direction are both chosen from predefined ranges, [*Vmin;Vmax*] and $[0;2\pi]$ respectively. Each movement in the Random Waypoint Mobility Model occurs in either a constant time interval *t* or after a constant distance traveled *d*, after which there is a pause *t*, and at the end of which a new direction and speed are calculated and changed. To determine the position coordinates of any node at any point of time we use the following calculations: Suppose initially a node is at any point (x_0, y_0). It travels for time *t* seconds, at a random angle θ_1 with velocity v_1, after which it again changes direction. The new position (x_1, y_1) of the Mobile Node (MN) will now be:

$$\mathrm{x_1 = x_0 + v_1 \cdot t \cdot \cos\theta_1} \tag{1}$$

$$\mathrm{y_1 = y_0 + v_1 \cdot t \cdot sine\,\theta_1} \tag{2}$$

We continue this position calculation process to track further movements of the MN within the network. So at any time *t* the position of the device is determined by the formula:

$$\mathrm{x_n = x_{n-1} + v_n \cdot t \cdot \cos\theta_n} \tag{3}$$

$$\mathrm{y_n = y_{n-1} + v_n \cdot t \cdot sine\,\theta_n} \tag{4}$$

Considering a network, where there are multiple leader nodes, service providers and clients,

1. The threshold value t is calculated and stored by each leader node in the network, based on battery life (b) of the leader node, the number of nodes registered with the leader node (n), and the distance from a registered node (d), as:

$$\mathrm{t = w_1 \cdot b(w_2 \cdot n\, w_3 \cdot d)} \tag{5}$$

2. If a new leader node enters into the network, then if it finds any other leader node within range, it registers with it. The new leader node also allows any unregistered service provider or client node to register with it.

3. If a new service provider or client node joins the network, then it searches for a leader node to register with. If there are multiple nodes within range, then it selects the leader node with the minimum positive threshold value.
4. In case a leader node moves out of range, then the nodes that were registered with it follow Step 3 to register with a new leader node.

Each leader stores its threshold value, which is recalculated and updated every time a new node registers with it. The threshold value is calculated as:

$$\mathrm{t} = \mathrm{f}(\mathrm{b}, \mathrm{n}, \mathrm{d}) \tag{6}$$

$$\text{i.e.}\quad \mathrm{t} = \mathrm{w}_1 \cdot \mathrm{b}(\mathrm{w}_2 \cdot \mathrm{n}\, \mathrm{w}_3 \cdot \mathrm{d}) \tag{7}$$

where w_1, w_2 and w_3 are the respective weights to be assigned to the corresponding parameters. Here all the weights are assumed to be 1. The following calculations are used to determine the threshold value of a leader node:

1. For determining the battery life '*b*', of each leader node, four values are considered:

Value	Range of remaining battery life
1	If the remaining battery life is >80 % of total capacity
2	If the remaining battery life is within 60–79 %
3	If the remaining battery life is within 30–59 %
−1	If the remaining battery power is <30 %

   To avoid registration with a node which has a remaining battery life of less than 30 %, the value of *b* and the function t have been chosen in such a way that it should lead to a negative threshold value if the battery life goes below 30 %.

2. The machine load '*n*' is denoted to be the number of nodes registered with a leader node at any point of time. So it is considered that $n =$ number of service provider or client nodes registered with a leader node at any point of time.
3. The distance of each leader node from a service provider/client is denoted as '*d*' and is measured using the standard Euclidean method to calculate the distance between any two points. The calculations of the positions of the service provider are done using the Random Way-point Mobility Model, as mentioned earlier in this section.

After the initial deployment, every subsequent entry or exit of the participating nodes that occurs will trigger the calculation of the threshold value of all the current leader nodes. Depending on this threshold value the service provider and client nodes will choose to register itself on a leader node among the multiple neighboring leader nodes. Algorithm 1 is the proposed algorithm.

```
for each node type = “leader node” {
    if new node registered then update {
        b← remaining battery life
            n ← number of registered nodes
            d ← distance from sp or client node
            threshold value $t = b.w1(n.w2+d.w3)$ } }
        Case 1: If (node entry= “service provider”) then {
          Receive multicast message to discover new
          lookup service within its neighbor
          If more than one leader node within range then
            find leader node with minimum positive threshold value t and register with it } }
        Case 2: If (node entry = “client”) then {
          Search for a lookup service.
            If more than one leader node within range,
            Then {find leader node with minimum positive
              threshold value and register with it } }
        Case 3: If (node entry = “leader node”) then {
          if within range of another leader node, then {
            registers with it then {
                allows unregistered service providers to
                register on it }}}
        Case 4: If (node exit = “leader node”) then {
        Service providers registered with this leader node will
        start to discover another leader node;
        Receive multicast message to discover new lookup service within its neighbor;
        if more than one leader node within range, then {
            register with leader node with minimum positive threshold value}}}
```

5 The Network Wide Search

Finally in this section, we present the process of the network wide search that a client performs when it searches for a service. The dynamic and self-configuring nature of Mobile Ad Hoc networks lead to formation of arbitrary topologies in the network [8]. The nodes are free to move randomly and organize themselves arbitrarily; thus, the network’s wireless topology may change rapidly and unpredictably. Here the service provider sends a multicast registration request to every other node in the network to discover leader nodes. If it finds one or more such leader node, then it registers on a suitable leader node based on the minimum threshold value as described in the Leader Node Selection algorithm in Sect. 5. Now when the client needs to avail a particular service, it needs the service proxies of the registered services from the lookup service of the leader node on which that service provider providing the service is registered. But the client may not be within the range of communication of the leader node on which the required service is available. So a network-wide search for the service is performed here by utililizing the multi-hop nature of MANETs. The client sends its search request to the leader node within its direct communication range and with which it is registered. If the service is not available on the lookup service of this leader node, then this request is propagated by the concerned leader node to other leader nodes in the network as well using multi- hop. In Jini the client search request initiates with the default *ttl* (time to live) value 15. Each propagation of this search request by a node will decrease the *ttl* value by one. When the *ttl* value becomes 0 the search request will not proceed any further. This default *ttl* value can be changed by setting the appropriate property in the Jini system. The client search request may match with the service proxies on multiple lookup services. All these matched service proxies on the lookup services are unmarshalled by the client for further use. Figure 1 represents network wide searching in this system.

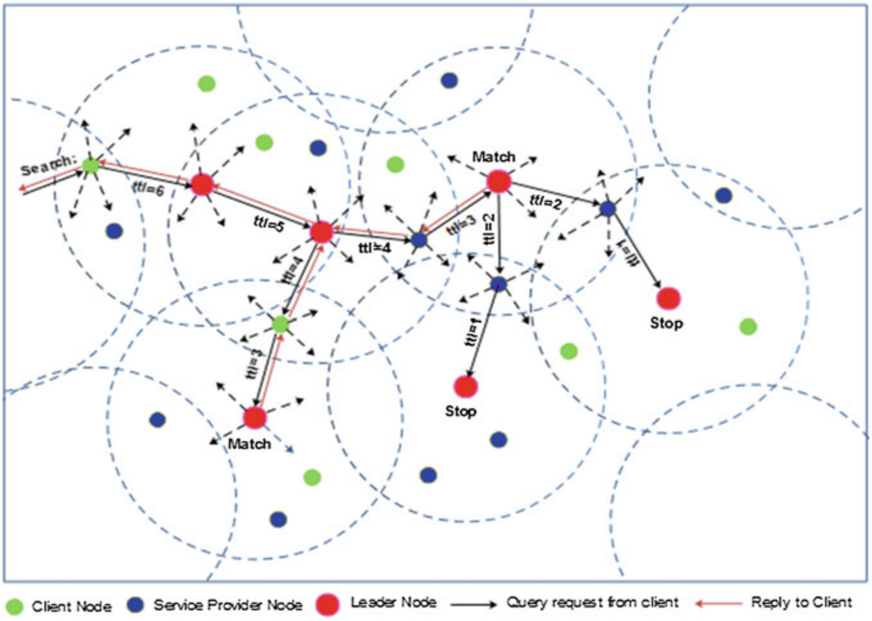


Fig. 1 The network wide searching

6 Conclusion and Future Work

This application has been tested within a limited range using 7–8 nodes. In future we need to test this framework on a larger real time area and with the number of nodes scaled up. However to be implemented on a larger scale, the implementing agencies need to follow uniformity in the design of the GUIs and other compatibility issues need to be smoothened out. The framework can be adapted to other scenarios also with minor modifications in the fields. Additionally the weights of all the parameters for determining node suitability have been taken to be 1. In future the weights need to be fine tuned and set to get more accurate results.

References

1. Kevin B, Kevin M, Scott R (2003) Self-adaptive leasing for JINI. In: PerCom 2003:539–542
2. Zhu F, Mutka MW, Ni LM (2005) Service discovery in pervasive computing environment. IEEE Pervasive Comput 4:81–90
3. Muhammad MR, Abdullah-Al-Wadud M, Chae O (2008) Performance analysis of leader election algorithms in mobile ad hoc networks. Int J Comput Sci Netw Secur 8(2):379–388
4. Zhang G, Kuang X, Chen J, Yu Z (2009) Design and implementation of a leader election algorithm in hierarchy mobile ad hoc network. In: Proceedings of 4th international conference on computer, science and education
5. Lee SS, Muhammad RM, Kim CG (2007) A leader election algorithm within candidates on ad hoc mobile networks. In: Embedded software and systems: third international conference, ICESS 2007
6. Roy M, Kar P, Mukherjee N (2010) An enhanced service framework in MANETs for application in emergency services. In: Proceedings of international conference on intelligent network and computing
7. Camp T, Boleng J, Davies V (2002) A survey of mobility models for ad hoc network research. Wirel Commun Mobile Comput (WCMC): Special Issue Mobile. Ad Hoc Netw 2(5):483–02
8. Roy M, Kar P, Mukherjee N (2010) Determining JINI leasing time limits using the random waypoint mobility model in mobile ad hoc networks. In: Proceedings of international conference on networking and information technology

A Novel Methodology for Securing Ad Hoc Network by Friendly Group Model

Md. Amir Khusru Akhtar and G. Sahoo

Abstract MANETs may be considered as a society in which nodes agree to cooperate with each other to fulfill the common goal. But noncooperation is genuine to save itself in terms of their battery power and bandwidth. Ad hoc network is still a challenge as lots of work has been proposed but they have serious limitations in terms of routing overhead and attacks. Modification of routing information's can be handled by secure routing protocols but non cooperation is still in its natal stage. Our proposed FG Model minimizes resource utilization by cutting down the routing overhead, so that less number of nodes participates in routing activities by creating smaller friendly groups. The proposed method minimizes routing overhead to a ratio of $k\colon 1|k > 1$ (where k is the no. of friendly groups) and it can be implemented on the existing MNAET routing protocol (e.g., DSR, AODV and TORA).

Keywords Open MANET · Closed MANET · Friendly groups · Cooperative society · Border node · Regular node · Gateway

1 Introduction

Ad hoc wireless networks are infrastructure-less, self-organized network that can be created in minimum time. Cooperation is the backbone of such type of network because they have surplus responsibilities such as routing & addressing. In spite of

M. A. K. Akhtar (✉)
Department of Computer Science and Engineering,
ICFAI University, Ranchi, Jharkhand, India
e-mail: akru2008@gmail.com

G. Sahoo
Department of Information Technology, Birla Institute of Technology,
Mesra, Ranchi, Jharkhand, India
e-mail: gsahoo@bitmesra.ac.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_3,

that nodes are free to move in the network without loosing connection. MANETS are difficult to implement due to their distributed and dynamic nature. A Protection model characteristic is not only to protect the existing attacks (malicious and selfish) but it should also minimize the routing overhead. MANET is the only option for many applications such as Military and Law, Disaster relief operations, Mine site operations and other suitable domain when infrastructures are not available, impractical, or expensive.

A MANET is a network of cooperation but when a node showing its selfishness all cooperation agreement fails. Selfishness is natural behavior and it must not be denied, it is genuine. These honest causes (i.e., to save battery life, to save bandwidth) encourages nodes to become selfish. Existing MANET protocols can prevent to some extent but with a serious overhead that is sometimes similar to the cost if nodes are not selfish. It can't be prevented by rewarding [1] or by enforcing some complex calculation [2–5]. Defining some new way to confront from this problem, it in terms of minimizing the routing activities that saves battery life, so that the non ethical behavior will not take place up to the maximum extent.

The motivation behind our approach is that network partitioning can improve the overall network throughput that solves both the malicious and selfish attacks. In this work we are categorizing MANETs into closed MANET and open MANET. Closed MANETs works together for a common goal and therefore selfishness is not often expected because they have some defined objective as in Military or police exercises, Disaster relief operations, and Mine site operations. On the other hand open MANETs are formed for diverse goals, they agree to share resource but for saving itself they may become selfish even though they don't like this.

MANETs issues a new future for Business meetings, home networking or distributive communication and computing. Designing a complex prevention scheme or defining some detection and exclusion mechanism for discouraging selfish behavior is not sufficient because still it consumes battery power and available bandwidth that is the real cause of selfishness.

Our proposed FG model divides a MANET of size N into k friendly groups with approx N/k number of nodes, which minimizes the throughput up to a ratio $k\colon 1|k > 1$ where k represents number of groups. It consumes less battery power and bandwidth by minimizing the routing overhead this will establish a cooperative society for the successful execution of MANET.

The approach of this paper is organized as follows. Section 2 enlighten the related works in the field of selfish node prevention or detection and exclusion, and the impact of physical and virtual subnetting or partitioning in minimizing total routing overhead. Section 3 discusses the FG Model. Experimental analysis is given in Sect. 4. Section 5 addresses efficiency evaluation for reactive routing protocol with and without FG model. Finally Sect. 6 concludes the paper.

2 Related Work

Ad hoc network is still a challenge as lots of work has been proposed but they have serious limitations in terms of routing overhead and attacks. Modification of routing information's can be handled by secure routing protocols [6–10] but non cooperation is still a challenge for the existing secured routing protocols. A variety of routing protocols have been proposed but selfishness is still in its natal stage.

Works done on the area of detection of misbehavior using Reputation based mechanism are as follows.

"Watchdog" and "Pathrater" [1] mechanism was proposed to be used over the DSR [11] routing protocol but selfish nodes are rewarded because there is no punishment for the same. It has another serious drawback that extra battery power consumption because every node has to constantly listen to the medium.

CONFIDANT, "Cooperation of Nodes: Fairness In Dynamic Ad-hoc Networks" [2]. It has four components (a monitor, a reputation system, a path manager, and a trust manager) to achieve cooperation between nodes, but these four components are implemented in every node, as it creates lots of overhead.

CORE, "a collaborative reputation mechanism" [3], it uses Watchdog for monitoring mechanism with a reputation table. This mechanism punishes the selfish nodes but it is also very costly.

Friends & Foes [4] in which friend receives the services and Foes that is refuses to serve by the nodes. This method provide solution but with memory and message overhead.

RIW [5] this method gives emphasis is given on current feedback items rather than old ones. It keeps a node behaving selfishly for a long time after building up good reputation, but with impractical assumption.

Works done on the area of detection of misbehavior using Incentive based mechanism are as follows.

PPM/PTM [12] it uses two models the Packet Purse Model that loads Nuglets into data packets before sending them for the payment to intermediate nodes. Intermediate nodes can take more Nuglets than they deserve, and Packet Trade Model that maintains intermediate nodes trade packets with the previous node, and the destination finally pays the price of the packets. This model needs a secure hardware to keep nodes from tampering the amount of Nuglets.

Ad hoc VCG (Vickery, Clarke and Groves) [13] it has two phases the Route Discovery in which destination node computes needed payments for intermediate nodes and notifies it to the source node or the central bank, and Data Transmission phase where actual payment is performed. In VCG nodes totally depends on destination nodes report.

Sprite [14] works on Central Authorized Server (CCS). Nodes have to send a receipt to CCS for every packet they forward; CCS assigns credits to nodes according to the receipt. Scalability and message overhead are the major weaknesses.

Priority forwarding [15] uses two layered forwarding: Priced Priority forwarding and Free best-effort forwarding but it has packets forwarding problems.

PIFA (Protocol independent Fairness Algorithm) [16] is suitable for any routing protocol. It introduces Credit Manager (CM) that manages the credit databases for nodes, Bank Stations or sink nodes. Message processing overhead is the major weakness.

Game theory based schemes [17–19] and other schemes [20, 21] have weaknesses in terms of routing and processing overheads.

A Partition Network Model was proposed to minimize the routing overhead with Mobile agents [22]. This method minimizes overhead but not suitable for many applications.

For the minimization of routing overhead subnetting concepts was proposed [23] that uses a internet type structure in which the nodes are grouped into subnets acting as a single entity but it is difficult to apply due to their dynamic and distributed nature. It includes open challenges such as subnet formation and address acquisition, the intra-subnet routing and inter-subnet routing, and the mobility of nodes between subnets.

Some of the papers proposed efficient Virtual subnet model for MANET [24–26] but it is not suitable for devices having low computation power because each node in the subnet is authenticated using certificate and it involve so many computations.

3 Friendly Group Model (FG)

3.1 Overview

The proposed Friendly Group Model (FG) divides a MANET into a number of friendly groups, with one border group. Each friendly group consisting of regular nodes with one border node and it should be a member of the border group. The FGs are defined on the basis of common objectives.

Figure 1 shows our Friendly Group structure. The four different FGs are represented by cross, triangle, circle and square. In a FG Regular nodes are responsible for intra group routing while the border nodes are responsible for intra and inter group routing.

3.2 Elementary Terminologies

Definition 1. *Regular Node (RN): it represents a static or mobile terminal performs both terminal and routing functions within its friendly group.*

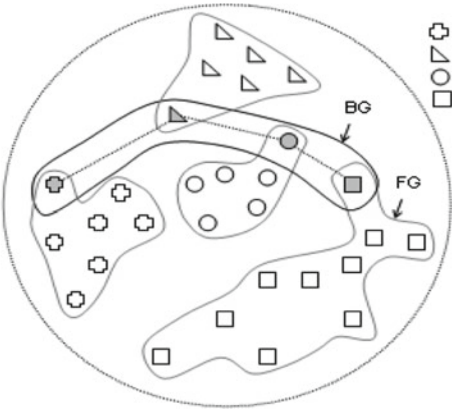


Fig. 1 Friendly group architecture of four FG with one BG

Definition 2. *Border Node* (*BN*): *it represents a static terminal performs terminal and routing functions within its friendly group and in the border group*.

Definition 3. *Closed MANET* (*CM*): *it represents a MANET containing set of nodes works together for common objectives*.

Definition 4. *OPEN MANET* (*OM*): *it represents a MANET containing set of nodes formed for diverse goals*.

Definition 5. *Friendly Group* (*FG*): *it represents a set of regular nodes together with one border node which has common objectives* (*Closed MANET*).

Definition 6. Border *Group* (*FG*): *it represents a set of border nodes used as gateways for the FGs which has diverse objectives* (*Open MANET*).

3.3 FG Components

Our approach involves the following FG components.

- Ad hoc Component: It includes ad hoc related protocols installed in nodes and connected with ad hoc networks.
- The border node must be equipped with two wireless devices to support multiple networks and similar concept was defined in [27]. The first NIC connected to the FG and the second NIC is connected to the BG as given in Fig. 2.

3.4 Group Formation

Some assumption has been taken for the grouping. To make a group a set of nodes form a FG in which nodes having common goal. For example employees of several

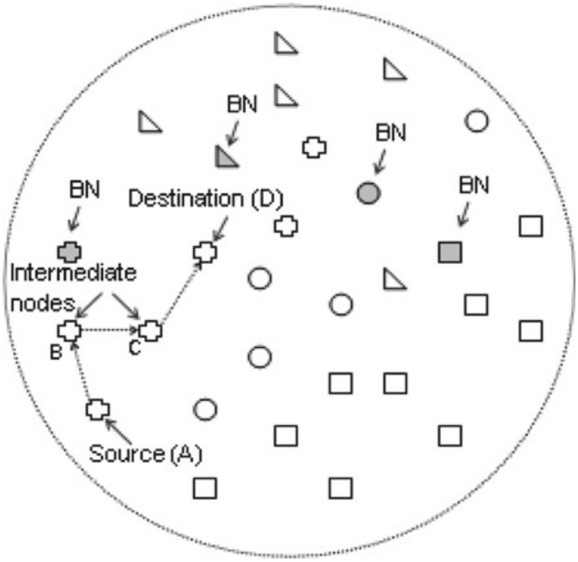


Fig. 2 Within group transmission

departments can be grouped into common objective groups like marketing (FG1), finance (FG2), HRD (FG3) and Personnel (FG4).

3.5 *Transmission of Packets*

3.5.1 Within a Friendly Group (Closed MANET)

Packet addressed to the same FG is forwarded to the destination using existing MANET routing protocol (e.g., DSR, AODV, and TORA) [11, 28, 29]. The intra group routing is given in Fig. 3.

Here node A is the source and node D is destination. The dotted arrow shows the data flow and the path is A → B → C → D. The routing operation will be performed without the help of BN. However, while routing, if BN node of same group participates in routing activities, then BN node would act like a RN.

3.5.2 Inter Group (Open MANET)

Packet addressed to a destination of other FG routed through the border node. The borders nodes are the gateway for FGs. BNs use MANET routing protocols (e.g., DSR,

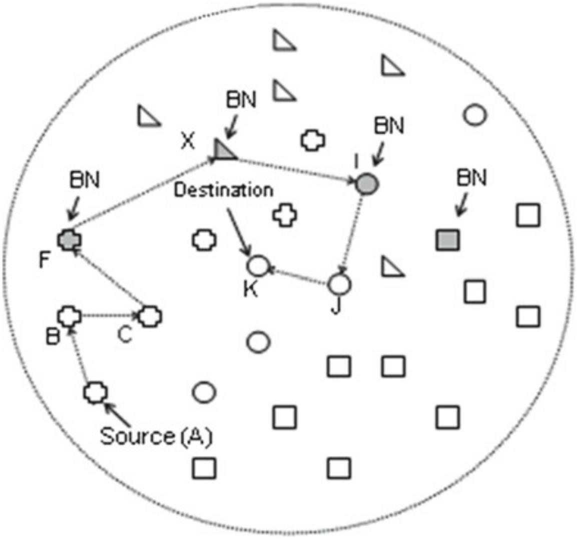


Fig. 3 Inter group transmission

AODV, and TORA) to send and receive packets. The inter group routing is given in Fig. 4.

Here, the source node is a cross node A and destination node is a circle node K. The dotted arrow data flow shows that packet is first forwarded to BN of the cross FG

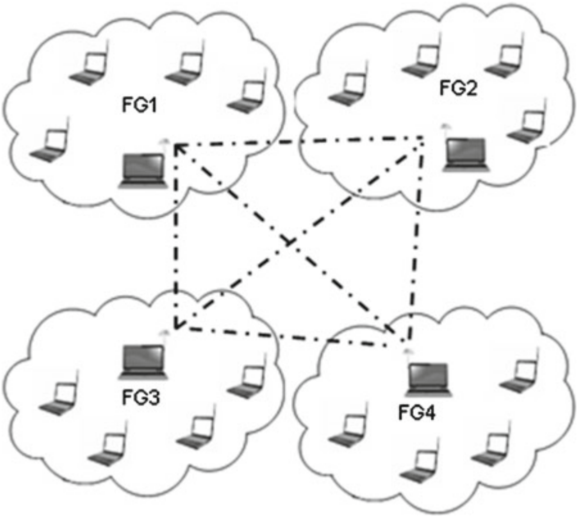


Fig. 4 Experiment overview

with path A → B → C → F. After receiving the packet the BN checks the packet to identify destination. If the destination is on other FG then border group routing will be performed to sends the packet to the destination. The complete flow from source to destination is shown by the dotted arrow and the packet travels the path A → B → C → F → X → I → J → K.

4 Experiments on Lab

4.1 Scenario Description

Figure 5 shows a scenario in which our proposal is based. We have deployed a Friendly group structure, over the Lab with 20 nodes. We select 4 nodes as the border nodes. RNs having single NIC and BNs are equipped with two NICs.

4.2 Regular Node Configuration

The RNs containing single NIC and it is connected to FG as given Fig. 6.

The RNs default gateway will be used when a destination IP address does not fall within the FG, and it is used to route packets to other group.

4.3 Border Node Configuration

The BN multiple NIC configurations are given in Fig. 2. The Border node is connected to a FG via NIC1 which allows routing and forwarding within group and NIC2 is connected to BG to provide inter routing.

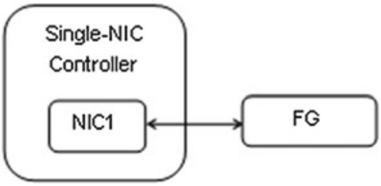


Fig. 5 RNs having single NIC controller

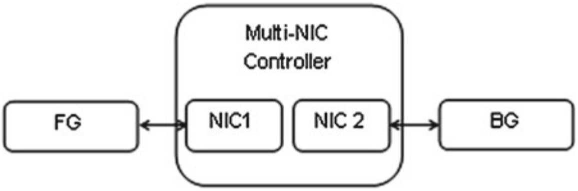
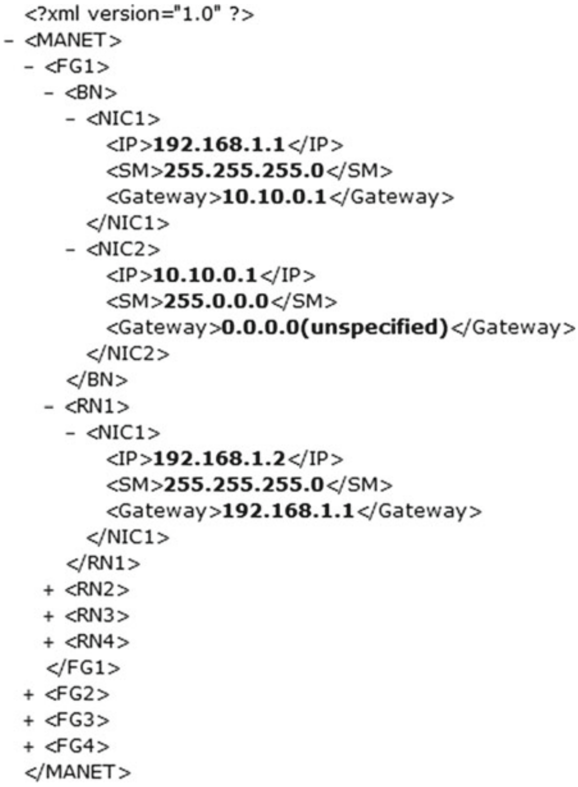


Fig. 6 BNs having multiple NIC controller

4.4 IP Structure

The FGs IP addresses in the range of 192.168.x.x with subnet mask 255.255.255.0 to NIC1. The default gateway address is 10.0.0.x for FGs. In this FG Model, only the NIC1 default gateway will be used because when a destination IP address does not fall within the FG of any NIC, then the default gateway of NIC1 is used to route packets to other group. The NIC2 default gateway should be blank because it routs the packets to the wrong network and causing them to be undeliverable.

The IP structure for FG Model is shown in Fig. 7 in which IP denotes internet protocol address and SM denotes Subnet Mask.



```
  <?xml version="1.0" ?>
- <MANET>
  - <FG1>
    - <BN>
      - <NIC1>
          <IP>192.168.1.1</IP>
          <SM>255.255.255.0</SM>
          <Gateway>10.10.0.1</Gateway>
        </NIC1>
      - <NIC2>
          <IP>10.10.0.1</IP>
          <SM>255.0.0.0</SM>
          <Gateway>0.0.0.0(unspecified)</Gateway>
        </NIC2>
      </BN>
    - <RN1>
      - <NIC1>
          <IP>192.168.1.2</IP>
          <SM>255.255.255.0</SM>
          <Gateway>192.168.1.1</Gateway>
        </NIC1>
      </RN1>
    + <RN2>
    + <RN3>
    + <RN4>
    </FG1>
  + <FG2>
  + <FG3>
  + <FG4>
  </MANET>
```

Fig. 7 FG IP structure

4.5 Shifting of Node

Due to the dynamic nature of MANET the mobility of the nodes into the friendly group can be handled by MANET routing protocol. The mobile nodes need to discover the border node which is in the best coverage area of the FG.

5 Efficiency Evaluation

5.1 Overhead Comparisons

We have analyzed the overhead for reactive routing protocols [30], and the proposed subnetting concept [23] for a MANET that reduces routing overhead with some open challenges such as subnet formation and address acquisition, the intra-subnet routing and inter-subnet routing, and the mobility of nodes between subnets.

Our FG model is very simple because it can be implemented within a short time and it does not involve any further computations to enforce authentication. This model classify MANETs into open MANETs and closed MANETs and works efficiently using existing routing protocols. The classification minimizes total no. of control packets in routing activities.

The reduction of the routing information is achieved by partitioning the network into FGs and the FGs are joined using BG. This filtered the control packet overhead and our protocol assumes that inter group routing will happen rarely.

The overhead comparison between a reactive routing protocol and the same protocol with friendly Groups is given in Fig. 8. The result shows the reduction in control packets by introducing FGs. Our model divides a MANET of size N into k friendly groups with approx N/k number of nodes per group. The total control overhead in is N^2 with flat structure and in FG (Hierarchical) structure it N^2/k which minimizes the throughput up to a ratio k: $1|k > 1$.

5.2 Advantage of FG Model

This model uses two NIC instead of virtual subnetting [24–26] because of the following advantages:

- Simpler to design because it divides a MANET by involving NICs
- Cost effective because dividing a network by employing NICs is very cheaper
- Negligible performance overheads due to its design but in virtual subnetting nodes are authenticated using certificate that involve so many computations.
- Suitable for devices having low computation power because it does not involve any computations for authentication.

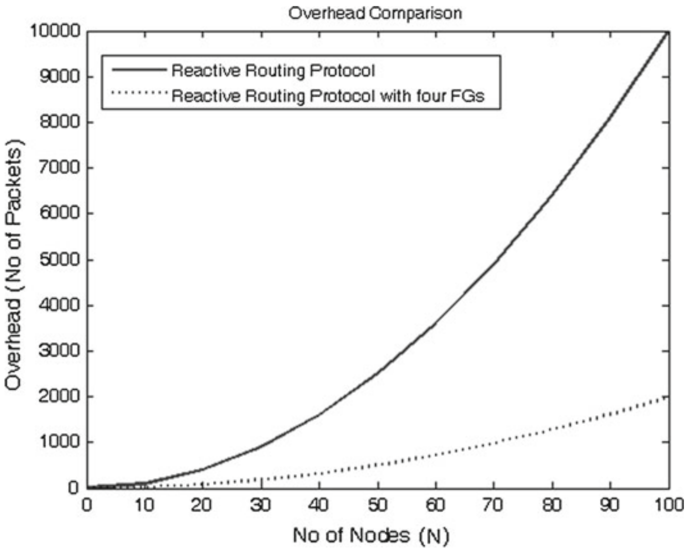


Fig. 8 overhead comparison between a reactive routing protocol and the same protocol with friendly Groups

6 Conclusion

Our Friendly Group Model for MANET will encourage nodes cooperation because it reduces the routing overhead. This model divides a MANET of size N into k friendly groups with approx N/k number of nodes, which minimizes the throughput up to a ratio k: $1|k > 1$ where k represents number of groups. This model saves battery life and bandwidth and thus enhances cooperation by eliminating the genuine cause of selfishness.

This model assumed that by saving battery power nodes have fewer chances for misbehaving and it is true also because they have enough energy to survive.

References

1. Marti S, Giuli TJ, Lai K, Baker M (2000) Mitigating routing misbehavior in mobile ad hoc networks. In: Proceedings of the 6th annual international conference on mobile computing and networking (ACM MobiCom 2000). ACM, New York, pp 255–265
2. Buchegger S, Boudec J-YL (2002) Performance analysis of the CONFIDANT protocol: cooperation of nodes—fairness in distributed ad-hoc networks. In: MOBIHOC'02
3. Michiardi P, Molva R (2002) CORE: a collaborative reputation mechanism to enforce cooperation in mobile ad-hoc networks. In: CMS'2002, communication and multimedia security 2002 conference, 26–27 Sept 2002, Portoroz, Slovenia/Also published in the book : Advanced communications and multimedia security, Jerman-Blazic B, Klobucar T (eds), Kluwer Academic Publishers, Dordrecht, ISBN 1-4020-7206-6, Aug 2002, 320 pp

4. Miranda H, Rodrigues L (2003) Friends and foes: preventing selfishness in open mobile ad hoc networks. In: ICDCSW'03
5. Adams WJ, Hadjichristofi GC, Davis NJ IV (2005) Calculating a node's reputation in a mobile ad hoc network. In: Proceedings of IEEE international performance computing and communications conference (IPCCC), pp 303–307
6. Sanzgiri K, Dahill B, Levine B, Shields C, Belding-Royer E (2002) A secure routing protocol for ad hoc networks. In: 10th IEEE international conference on network protocols (ICNP), Nov 2002
7. Zapata MG, Asokan N (2002) SAODV: securing ad-hoc routing protocols. In: 2002 ACM workshop on wireless security (WiSe 2002), Sept 2002, pp 1–10
8. Hu Y, Perrig A, Johnson D (2002) Ariadne: a secure on-demand routing protocol for ad hoc networks. In: Proceedings of the eighth annual international conference on mobile computing and networking, Sept 2002, pp 12–23
9. Hu Y, Johnson D, Perrig A (2002) SEAD: secure efficient distance vector routing in mobile wireless ad hoc networks. In: Fourth IEEE workshop on mobile computing systems and applications, June 2002, pp 3–13
10. Papadimitratos P, Haas Z, Samar P (2002) The secure routing protocol (SRP) for ad hoc networks, internet-draft, draft-papadimitratos-securerouting-protocol-00.txt, Dec 2002
11. Johnson D, Maltz D, Hu Y-C (2003) The dynamic source routing protocol for mobile ad hoc networks (DSR). IEEE internet draft, Apr 2003
12. Buttyan L, Hubaux J (2000) Enforcing service availability in mobile ad hoc WANs. In: Proceedings of IEEE/ACM MobiHOC workshop
13. Anderegg L, Eidenbenz S (2003) Ad hoc-VCG: a truthful and cost-efficient routing protocol for mobile ad hoc networks with selfish agents. In: Proceedings of ACM MobiCom 2003, pp 245–259
14. Zhong S, Chen J, Yang YR (April 2003) SPRITE: a simple, cheatproof, credit-based system for mobile ad-hoc networks. In: Proceedings of INFOCOM 03, pp 1987–1997, Apr 2003
15. Raghavan B, Snoeren AC (2003) Priority forwarding in adhoc networks with self-interested parties. In: Workshop on peer to peer systems, June 2003
16. Yoo Y, Ahn S, Agrawal DP (2005) A credit-payment scheme for packet forwarding fairness in mobile MANETs. In: Proceedings of IEEE ICC
17. Srinivasan V, Nuggehalli P, Chiasserini CF, Rao RR (2003) Cooperation in wireless MANETs. In: Proceedings of IEEE INFOCOM
18. Mahajan R, Rodrig M, Wetherall D, Zahorjan J (2005) Sustaining cooperation in multi-hop wireless networks. In: Proceedings of NSDI
19. Hales D (2004) From selfish nodes to cooperative networks—emergent link-based incentives in peer-to-peer networks. In: Proceedings of IEEE international conference on peer-to-peer computing, pp. 151–158
20. Yang H, Meng X, Lu S (2002) Self-organized network-layer security in mobile ad hoc networks. In: Proceedings of the ACM workshop on wireless security. ACM, New York, pp 11–20
21. Feigenbaum J, Papadimitriou C, Sami R, Shenker S (2002) A BGP based mechanism for lowest-cost routing. In: Proceedings of the 21st annual symposium on principles of distributed computing. ACM, New York, pp 173–182
22. Chiang T-C, Tsai H-M, Huang Y-M (2005) A partition network model for ad hoc networks. In: Wireless and mobile computing, networking and communications (WiMob'2005), pp 467–472
23. López J, Barceló JM, García-Vidal J (2006) Subnet formation and address allocation approach for a routing with subnets scheme in MANETs. In: Wireless systems and network architectures in next generation internet. Lecture notes in computer science, vol 3883/2006, pp 62–77. DOI:10.1007/11750673_6
24. Chowdhury MAH, Ikram M, Kim K-H (2008) Secure and survivable group communication over MANET using CRTDH based on a virtual subnet model. In: IEEE Asia-Pacific services computing conference
25. Chang C-W, Yeh C-H, Tsai C-D (2010) An efficient authentication protocol for virtual subnets on mobile ad hoc networks. In: International symposium on computer, communication, control and automation

26. Vilhekar AA, Jaidhar CD (2012) Modified authentication protocol using C mobile adhoc networks. In: Wireless communications and applications. Lecture notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 72. Springer, Berlin, pp 426–432
27. National Instruments, www.ni.com/white-paper/12558/en. Accessed 06 June 2012
28. Perkins C, Royer EB, Das S (2003) Ad hoc on-demand distance vector (AODV) routing. IETF internet draft
29. Park V, Corson S (2001) Temporally ordered routing algorithm (TORA). IETF internet draft
30. Viennot L, Jacquet P, Clausen TH (2004) Analyzing control traffic overhead versus mobility and data traffic activity in mobile ad-hoc network protocols. Presented at ACM wireless networks journal (Winet)

Energy Efficient Medium Access Protocol for Clustered Wireless Sensor Networks

K. N. Shreenath and K. G. Srinivasa

Abstract Wireless sensor networks use battery-operated computing and sensing devices. A network of these devices will collaborate for a common application such as environmental monitoring. We expect sensor networks to be deployed in an ad hoc fashion, with individual nodes remaining largely inactive for long periods of time, but then becoming suddenly active when something is detected. These characteristics of sensor networks and applications motivate a MAC that is different from traditional wireless MACs such as IEEE 802.11 in almost every way: energy conservation and self-configuration are primary goals, while per-node fairness and latency are less important. In this proposal, we present a new MAC layer protocol for cluster based wireless sensor networks that reduces energy consumption and provides Quality of Service (QoS) through the use of service differentiation concept. The proposed protocol consists of two parts: First part is responsible for classifying gathered data at sensor nodes based on its importance and then stores it in the appropriate queue of the node's queuing system. The second part is responsible for energy efficient medium access mechanism that uses both scheduled and unscheduled schemes to gain a save in energy, and hence extending network's lifetime. The save in energy is achieved by differentiating between control and data messages. Data messages are assigned scheduled slots with no contention, while short control messages are assigned random access slots.

Keywords Sensor · Energy · Efficient · Scheduled · MAC · TDMA · Clustering

K. N. Shreenath (✉)
Department of CSE, Siddaganga Institute of Technology,
Tumkur, India
e-mail: shreenathk_n@yahoo.co.uk

K. G. Srinivasa (✉)
Department of CSE, M S Ramaiah Institute of Technology, Bangalore, India
e-mail: srinivasa.kg@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_4,

1 Introduction

Wireless sensor networking is an emerging technology that has a wide range of potential applications including environment monitoring, smart spaces, medical systems and robotic exploration. Such a network normally consists of a large number of distributed nodes that organize themselves into a multi-hop wireless network. Each node has one or more sensors, embedded processors and low-power radios, and is normally battery operated. Typically, these nodes coordinate to perform a common task.

Medium access control (MAC) is an important technique that enables the successful operation of the network. One fundamental task of the MAC protocol is to avoid collisions so that two interfering nodes do not transmit at the same time. There are many MAC protocols that have been developed for wireless voice and data communication networks. Typical examples include the time division multiple access (TDMA), code division multiple access (CDMA), and contention-based protocols like IEEE 802.11.

A good MAC protocol should have the following attributes. The first is the energy efficiency. Sensor nodes are likely to be battery powered, and it is often very difficult to change or recharge batteries for these nodes. Prolonging network lifetime for these nodes is a critical issue. Another important attribute is the scalability to the change in network size, node density and topology. Some nodes may die over time; some new nodes may join later; some nodes may move to different locations. The network topology changes over time. A good MAC protocol should easily accommodate such network changes.

Energy constraints of sensor networks have demanded energy awareness at most layers of the networking protocol stack. Provided that the radio transceiver unit considered as the major consumer of energy resource of the sensor node especially when the radio transceiver is turned on all time, then a large amount of energy savings can be achieved through energy efficient media access control mechanisms. For this reason, energy consideration has dominated most of the research at MAC layer level in wireless sensor networks. The concepts of latency, throughput and delay were not primary concerns in most of the presented work on sensor networks.

We propose a new energy efficient medium access control mechanism with quality of service support for cluster based wireless sensor networks. It uses a both scheduled (TDMA) and contention based (CSMA) medium access schemes. It differentiates between short and long messages; long data messages are assigned scheduled TDMA slots, while short periodic control messages are assigned random access slots. This technique limits message collisions and reduces the total energy consumed by the radio transceiver. Perhaps the greatest advantage is the efficient node's battery usage and its support for quality of service based on the service differentiation concept. The service differentiation is done through data prioritization to distinguish between four different priority levels based on traffic importance and criticality, while maintaining energy efficient usage of the sensor node's battery. It handles highest priority data packets differently than those of lowest priority. If the packet to be delivered is an urgent data packet that should be processed immediately without any delay, then the sensor node puts this packet into its high priority queue of its queuing system. This allows sensor nodes to do some type of traffic management and provide extremely highest priority traffic a greater chance of acquiring the channel and hence rapidly served with minimum delay.

2 Related Work

The medium access control protocols for the sensor networks can be classified broadly into two categories: Contention based and Schedule based.

2.1 Contention Based

The contention based protocols relax time synchronization requirements and can easily adjust to the topology changes as some new nodes may join and others may die few years after deployment. These protocols are based on Carrier Sense Multiple Access (CSMA) technique and have higher costs for message collisions, overhearing and idle listening.

The IEEE 802.11 [1] is a well-known contention based medium access control protocol which uses carrier sensing and randomized back-offs to avoid collisions of the data packets. The Power Save Mode of the IEEE 802.11 protocol reduces the idle listening by periodically entering into the sleep state. This PSM mode is for the single-hop network where the time synchronization is simple and may not be suitable for multi-hop networks because of the problems in clock synchronization, neighbor discovery and network partitioning.

Power Aware Multi-Access [2] is one of the earliest contention based MAC protocol designed with energy efficiency as the main objective. In this protocol nodes which are not transmitting or receiving are turned off in order to conserve energy. This protocol uses two separate channels for the data and control packets. It requires the use of two radios in the different frequency bands at each sensor node leading to the increase in the sensors cost, size and design complexity. Moreover, there is significant power consumption because of excessive switching between sleep and wakeup states.

Sensor—MAC [3] a contention based MAC protocol is modification of IEEE 802.11 protocol specially designed for the wireless sensor network. In this medium access control protocol, the sensor node periodically goes to the fixed listen/sleep cycle. A time frame in S-MAC is divided into two parts: one for a listening session and the other for a sleeping session. Only for a listen period, sensor nodes are able to communicate with other nodes and send some control packets.

Timeout T-MAC [4] is the protocol based on the S-MAC protocol in which the Active period is preempted and the sensor goes to the sleep period if no activation event has occurred for a particular time period. The event can be reception of data, start of listen/sleep frame time etc. The energy consumption in the Timeout T- MAC protocol is less than the Sensor S-MAC protocol. But the Timeout T-MAC protocol has high latency as compared to the S-MAC protocol.

2.2 Scheduled Based

The schedule based protocol can avoid collisions, overhearing and idle listening by scheduling transmit and listen periods but have strict time synchronization requirements.

The traffic adaptive medium access (TRAMA) [5] is a Contention based protocol that has been designed for energy efficient collision free channel in WSNs. In this protocol the power consumption has been reduced by ensuring collision free transmission and by switching the nodes to low power idle state when they are not transmitting or receiving.

SMACS [6] is a schedule based medium access control protocol for the wireless sensor network. This MAC protocol uses a combination of TDMA and FDMA or CDMA for accessing the channel. In this protocol the time slots are wasted if the sensor node does not have data to be sent to the intended receivers. This is one of the drawbacks of this MAC scheme.

Low Energy Adaptive Clustering Hierarchy (LEACH) is a energy aware scheduled based MAC [7] protocol assumes the formation of clusters in the network. The cluster head manages each of the cluster sensor nodes. The cluster head collects the information from the other sensor nodes within its cluster, performs the data fusion, communicates with the other cluster head and finally sends the data to the control center. The cluster head performs the assignment of the time slots to the sensor nodes within its cluster. The cluster head inform the other nodes about the time slot when it should listen to other nodes and the time slot when it can transmit own data.

Hybrid MAC protocols combine the strengths of scheduled and unscheduled MAC protocols while compensating their weakness to build more efficient MAC schemes. Hybrid protocols use different techniques to conserve sensor battery power; some protocols differentiate between small and long data messages. Long data messages are assigned scheduled slots with no contention, whereas small periodic control messages are assigned random access slots. Other hybrid techniques adjust the behavior of MAC protocol between CSMA and TDMA depending on the level of the contention in the network. The greatest advantage of the hybrid MAC protocols comes from its easy and rapid adaptability to traffic conditions which can save a large amount of energy, but this advantage comes at the cost of the protocol overhead and complexity caused by the TDMA structure which limits the scalability and applicability range of the protocol.

In [8], authors propose Q-MAC scheme that attempts to minimize the energy consumption in a multi-hop wireless sensor network, while providing quality of service by differentiating network services based on priority levels. The priority levels reflect the criticality of data packets originating from different sensor nodes. The Q-MAC accomplishes its task through two steps, intra-node and inter-node scheduling. The intra-node scheduling scheme adopts a multi-queue architecture to classify data packets according to their application and MAC layer abstraction. Inter-node scheduling uses a modified version of MACAW [9] protocol to coordinate and schedule data transmissions among sensor nodes.

Our MAC uses a more energy efficient way to coordinate and schedule data transmissions among clustered sensor nodes by utilizing both scheduled and unscheduled schemes.

3 Protocol Description

Our MAC protocol is composed of two parts.

The first part uses a modified version of the queuing architecture of Q-MAC. It classifies packets based on their importance and stores them into the appropriate queue. The source node knows the degree of the importance of each data packet it is sending which can be translated into predefined priority levels. The application layer sets the required priority level for each data packet by appending two extra bits at the end of each data packet (Fig. 1). The number of bits used to distinguish priorities could be set according to the number of priority levels required.

The queuing architecture of the Q-MAC is composed of four queues (Fig. 2). Each packet is placed in one of the four queues—high, medium, normal, or low based on the assigned priority. During transmission, the MAC transmits according to the priority of the data packet.

The second part adapts scheduled and unscheduled schemes in an attempt to utilize the strengths of both mechanisms to gain a save in energy resources of the sensor node, and hence prolonging the lifetime of the sensor network. It provides scheduled slots with no contention (based on TDMA) for data messages and random access slots (based on CSMA/CA) for periodic control messages. In the design of protocol, we assume that the underlying synchronization protocol can provide nearly perfect synchronization, so that synchronization errors can be neglected.

MAC classifies sensor nodes within the sensor network into clusters. For clustering in the sensor network, we can use different types of algorithms like heuristics, weighted, hierarchical, grid schemes. For our MAC protocol we are using an hierarchical scheme called Energy Efficient Clustering Scheme (EECS) [10].

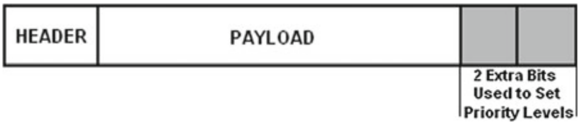


Fig. 1 Data packet format

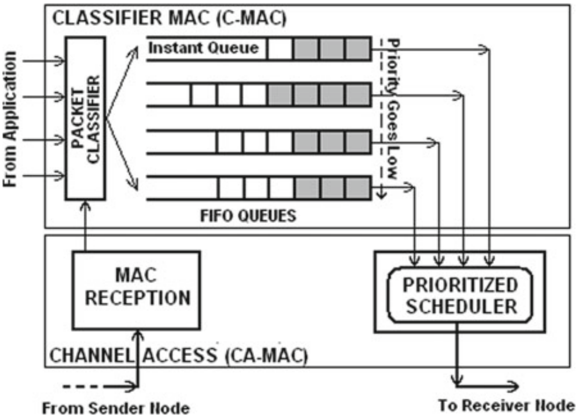


Fig. 2 Q-MAC structure

An EECS is a clustering algorithm in which cluster head candidates compete for the ability to elevate to cluster head for a given round. This competition involves candidates broadcasting their residual energy to neighboring candidates. If a given node does not find a node with more residual energy, it becomes a cluster head. Cluster formation is different than that of LEACH. LEACH forms clusters based on the minimum distance of nodes to their corresponding cluster head. EECS extends this algorithm by dynamic sizing of clusters based on cluster distance from the base station. The result is an algorithm that addresses the problem that clusters at a greater range from the base station require more energy for transmission than those that are closer. Ultimately, this improves the distribution of energy throughout of the network, resulting in better resource usage and extended network lifetime. The Cluster head node is responsible for controlling the channel access between sensor nodes and collects sensory data from them.

The transmission mechanism of MAC is based on dividing communication time into frames, which are controlled by the head node (Fig. 3). The frame is composed of two periods: contention period and normal period. Contention period is used to transmit and receive control signals, and consists of three parts; Frame Synchronization (SYNC), Request, and Receive Scheduling. Normal period is used to control the transmission of the gathered sensory data to head node.

Our proposed MAC accomplishes its task through the following four phases: Synchronization, Request, Receive Scheduling, and Data Transfer. Nodes that have data

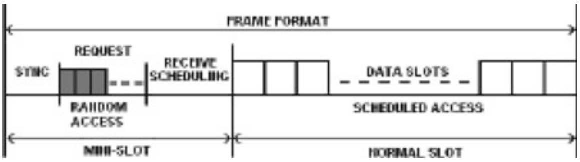


Fig. 3 Frame format of proposed MAC

to send should contend for the channel during the Request phase and send their requests along with the appropriate priority level of its traffic to the cluster head. Then, sensor nodes use the TDMA slots during the data transfer phase to send their data packets. The details are given below.

Synchronization phase: At the beginning of each frame, the head node broadcasts a SYNC message to all sensor nodes. All sensor nodes should be in receiving mode during this phase to be able to capture the SYNC message. The SYNC message contains synchronization information for the packet transmission.

Request phase: During this phase, sensor nodes that have data to transmit contend for the channel in order to acquire the access to send its request to the head node along with the required priority level. The Fig. 4 shows the request message structure.

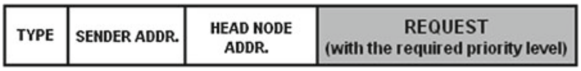
Receive Scheduling phase: The head node broadcasts a scheduling message to all sensor nodes that contains the TDMA slots for the subsequent phase data transfer phase. In this phase all sensor nodes should be in receive mode. Here, the cluster head should take into account the required priority levels when assigning TDMA slots to sensor nodes.

Data Transfer phase: In this phase, sensor nodes use the TDMA slots to transmit their data to the cluster head or to communicate with their neighbors. All sensor nodes that have no traffic to transmit or receive should turn their radio transceivers off and goes to sleep mode.

4 Performance Analysis

To study and evaluate the performance of the proposed MAC protocol, we have used NS2 sensor network simulator. The NS-2 simulation environment is a flexible tool for network engineers to investigate how various protocols perform with different configurations and topologies.

We evaluate the performance of the proposed protocol in a static single hop network topology. Energy consumption and average delay are the performance metrics used in the evaluation. The setup of experiments includes 100 nodes on a field of 100 × 100 m. The simulated nodes are within the radio range of each other. The simulation is allowed to run for 500 s and the results are averaged over several simulation runs. In simulation, we evaluate the performance of our proposed MAC and compare it with the standard S-MAC and Q-MAC. The simulation parameters are given in the Table 1.



TYPE	SENDER ADDR.	HEAD NODE ADDR.	REQUEST (with the required priority level)

Fig. 4 Request message format

Table 1 Simulation parameters

General		**S-MAC**	
Message payload	25 bytes	Frame length	610 ms
Data length	Up to 250 bytes	Contention window (CW)	15 ms
		Active period	300 ms
Radio			
Effective data rate	115 kbps	**Proposed MAC**	
Transmit	12 mA	Slots in SYNC phase (N1)	10
Receive	3.8 mA	Slots in scheduling phase (N2)	100
Sleep	0.7 μA	Slots in random access period (N3)	70
		Mini-slot time	1 ms
Q-MAC		Normal slot time	3 ms
Frame length	610 ms	Slots in data transfer phase (N4)	85
Contention period	300 ms	Frame length (N1+N2+N3)/3	+ N4
Contention window	15 ms		
Short space (SS)	0.5 ms		
Frame space (FS)	1 ms		

4.1 Energy Consumption

Energy efficiency is the most important performance metric in wireless sensor networks. The Fig. 5 shows the comparative energy consumption for the proposed MAC, S-MAC and Q-MAC. The proposed MAC protocol is more energy efficient than S-MAC and Q-MAC protocols. The proposed MAC protocol adapts better to the increase in the traffic rate and consumes less energy when compared to the other two protocols.

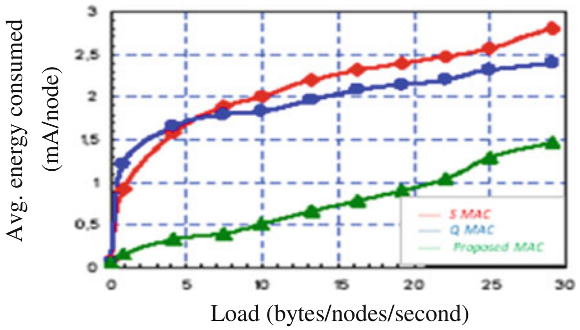


Fig. 5 Average energy consumption

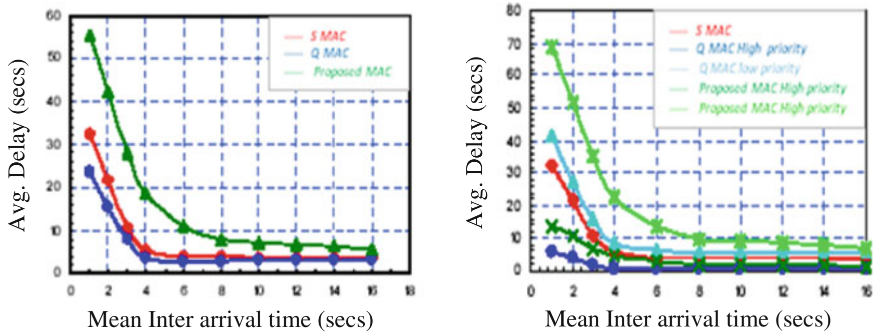


Fig. 6 Average packet delay (non prioritized and prioritized traffic)

4.2 *Average Packet Delay*

The Fig. 6 shows the average packet delay under normal and prioritized traffic. In the evaluation, we vary the traffic load by changing the packet inter-arrival time on the source node. The packet inter-arrival time changes from 1 to 16 sec. It is observed that the average delay time of contention based protocols are less than that of scheduled based protocols. This is because of the latency introduced by random scheduling. The results of prioritized traffic indicate that Proposed MAC successfully differentiates network services like Q-MAC. The higher priority packets are always accompanied with low latency. Therefore our MAC protocol achieves high energy efficiency under wide range of traffic loads and priorities.

5 Conclusion

We proposed a new energy efficient medium access control scheme for cluster based wireless sensor networks. The MAC combines the benefits of contention based and scheduled based protocols to achieve a significant amount of energy savings and offers QOS by differentiating network services based on priority levels. It enables only the nodes which have a data to transmit to access the channel according to their traffic priority levels; this avoids wasting slots by excluding those nodes which have no data to transmit from the TDMA schedule, and to switch nodes to sleep mode when they are not included in the communication process.

Prioritizing traffic according to its importance and criticality provides a greater chance for extremely highest priority nodes to access the channel and acquire the medium and hence rapidly served with minimum delay.

References

1. IEEE (1999) Wireless LAN medium access control (MAC) and physical layer specifications, 1999 edn. ANSI/IEEE Standard 802.11
2. Singh S, Raghavendra C (1998) PAMAS: power aware multi-access protocol with signaling for ad-hoc network. ACM SIGCOMM Computer Communication Review, July
3. Ye W, Heidemann J, Estrin D (2002) An energy-efficient MAC protocol for wireless sensor networks, IEEE INFOCOM, vol 2. New York, pp 1567–1576
4. van Dam T, Langendoen K (2003) An adaptive energy efficient MAC protocol for wireless networks. In: Proceedings of the first ACM conference on embedded networked sensorsystems. ACM Press, New York
5. Rajendran V, Obraczka K, Gracia-Luna-Aceves JJ (2003) Energy efficient, collision free medium access control for wireless sensor networks. In: ACM international conference on embedded networked sensor systems (SenSys), pp 181–192
6. Sohrabi K, Gao J, Ailawadhi V, Pottie GJ (2000) Protocols for self organization of a wireless sensor network. IEEE Pers Commun 7(5):16–27
7. Arisha K, Youssef M, Younis M (2002) Energy aware TDMA based MAC for sensor network. In: IEEE workshop on integrated management of power aware communications computing and networking (IMPACCT'02), New York
8. Yang L, Elhanany I, Hairong Q (2005) An energy-efficient QoS-aware media access control protocol for wireless sensor networks. Mobile adhoc and sensor systems conference
9. Bharghavan V et al (1994) MACAW: a media access protocol for wireless LANS. Proc. ACM SIGCOMM. 24(4)
10. Younis O, Fahny S (2008) Distributed clustering in Ad-hoc sensor networks: a hybrid energy efficient approach, IEEE Transaction on Mobile Computing, 2004, pp 248–340

A Design Mode of Streaming Media Transmission and Access in Wireless Video Sensor Network

Mengxi Xu, Chenrong Huang, Shengnan Zheng and Jianqiang Shi

Abstract Wireless video sensor network, acquired by urban road traffic flow information, is composed of several sensor network nodes with traffic flow information video detection, process and wireless communication capacity. Wireless digital camera sensor network cooperatively senses the traffic flow information (traffic flow, vehicle length, vehicle type, etc.) in each location by the means of each node. Network nodes transfer data to information convergence nodes (base station, SINK) by wireless multi-hop relaying mode. Meanwhile, adopting 2.4 G wireless communication, the convergence nodes analyze traffic flow video detection information and converge into MAN/WAN, Gigabit Ethernet, and the control center can do the query, record and other operations by reverse IPTV (interactive network television). This paper focuses on a design mode of streaming transmission and access application in wireless video sensor network.

Keywords Multimedia sensor network · Streaming media · Image communication · Intelligent traffic

M. Xu (✉)
School of Computer Science and Technology,
Nanjing University of Science and Technology,
Nanjing, China
e-mail: mengxi.xu@gmail.com

M. Xu · C. Huang · S. Zheng · J. Shi
School of Computer Engineering,
Nanjing Institute of Technology,
Nanjing, China

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_5,

1 Introduction

Logical information world is fused with real physical world by wireless sensor network, which deeply changes the interaction mode between human beings and natural. At abroad, WSN (Wireless Sensor Network) is originated from military application. From year 2002, WSN has been applied in environment, agriculture etc. and now it is expanded into monitoring, location and tracking, and environment sensor oriented intelligent computation in industry, electricity power, construction, intelligent traffic and so on. At present, one of the important aspects of WSN research is how to introduce image, audio, video and other media with abundant information into detection activities based on sensor network. Therefore, multimedia sensor network emerges as the time requires [1].

Recently, the research for multimedia sensor network technology has attracted scientific researchers. Some scholars do the exploratory researches on multimedia sensor network, meanwhile, they proposed some important research results in IEEE conferences (such as MASS, ICIP, WirelessCOM), ACM multimedia and sensor network conferences (such as ACM Multimedia, ACM MOBICOM, ACM WSNA). From year 2003, ACM specially organizes international video monitoring and sensor network seminar for communicating research results. University of California, Carnegie Mellon University, University of Massachusetts etc. has researched multimedia sensor network and set up video sensor network groups beginning with corresponding scientific plans. Meanwhile, domestic scholars also pay much attention to multimedia sensor network research.

In the research of multimedia sensor network, since the limitation of network resources, streaming media data need not to be transmitted in network in many applications, critical semantic information is abstracted to transmit by analyzing audio and video streaming data. Therefore, in one hand, the transmission burden of network is reduced, while the working lifetime can be extended, in the other hand, processing ability of nodes can be completely utilized to improve distribution of multimedia information processing in the whole system, which eases the burden of converging nodes, and improve the information processing rate of the whole system. Thus, intelligent multimedia information processing technology has an important affect for reducing network consumption, and improving the monitoring performance and quality. Two factors in intelligent multimedia information processing technology of multimedia sensor network need to be considered. One of the factors is complexity degree of processing, for the computational ability of multimedia sensor nodes is limited, it is not suitable for much complex processing technology; the other factor is the characteristics and application requirements of multimedia sensor network, traditional multimedia information processing technology needs to be improved to adapt multimedia sensor network.

In intelligent traffic application, traffic flow information includes: traffic flow, vehicle length classification, motorcade length, time occupied rate, space occupied rate, headstock distance, instant velocity, average velocity of time, average velocity of space and so on. Earlier traffic flow information detection mainly relied on labour count and investigation table, such as touch-tone labour count and inhabitant trip investigation table, which are still practical until now. In the end of the 1980', the detection technology of microwave radar, video image and light beacon began to be developed and utilized [2, 3].

Traffic flow video detection technology based on computer vision can take the place of traditional detectors and supply more status parameters of vehicles and traffic flow which can not be completed by other detectors. Traffic flow video detection technology is the key point researched both at home and abroad, which provides an effective method of traffic flow information acquisition for the development and practice of advanced traffic flow control system and intelligent vehicle system. The advantages of video detectors are: can be easily set up, can not damage the road surface, can not affect the traffic when constructing, can detect in a large area, the installation of the camera are convenient, flexible and easily maintained. The disadvantage is that it is limited by environmental factors such as image processing algorithm, atmosphere, light and shade. With the development of CCD technology and computer vision technology, video detectors can acquire large traffic flow information, which has a development potential.

Real-time traffic flow information acquisition is one of the weakest links in intelligent vehicle systems (IVHS) and advanced traffic management systems (ATMS). Present traffic flow detectors are insufficient in information acquisition, communication and reliability. With the development of CCD technology, computer vision and information processing technology, it can be fixed by wireless video detection. Wireless video detection has the abilities of more layout and wireless detection, meanwhile, it can acquire important traffic flow parameters of vehicle density, queue scale and park times, vehicle size which can hardly achieved by routine detectors. The development tendency of wireless video detection technology has an obvious advantages, good perspectiveness, and can represent the development tendency of traffic information detectors [4–7].

Wireless video detection technology occupies an important position in modern traffic systems, which is the development foundation of future intelligent vehicle systems. However, the present problems are low real-time, high error rate in image processing, and the detection accuracy is limited by software/hardware of the entire system. For utilizing the video image to acquire the complexity of the traffic flow information, wireless video detection technology is still continuously improved. Nevertheless, with the development of computer vision, image processing technology and microelectronics technology, wireless video detection technology needs to be continuously improved and widely applied.

2 Overall Design of Wireless Video Sensor Networks

Wireless video sensor network acquired by urban road traffic flow information is composed of several sensor network nodes with traffic flow information video detection, processing and wireless communication capacity. Wireless digital camera sensor network senses local traffic flow information (vehicle flow, vehicle velocity, vehicle length, vehicle types and so on) by the distributions of each node, while network nodes transmit data to information convergence nodes (base station, SINK) by wireless multiple hop relay. By 2.4 G wireless communication, traffic flow video detection information is analyzed by convergence nodes and converged to MAN/WAN and Gigabit Ethernet. Thus, master control center can do the query, record and other operations according to reversal IPTV. The architecture of wireless video sensor network acquired by traffic flow information is shown in Fig. 1.

The wireless video sensor network acquired by traffic flow information can realize IP broadband accessing network in the core of microwave transmission technology, networking conveniently, without excavating roads to interrupt traffic, which can achieve detection and management sharing the same wireless broadband network, being compatible with wired IP network and optical fiber, supporting multi-layer management integration, easily expanding intelligent wireless video system, and supporting one level to multilevel control center. Meanwhile, it can make each center detect multi-section multi video flows, call multi video flows of each section to conduct by grading and separation of powers. At the same time, it can operate safely and steadily in a large area, and easily maintain.

3 Data Management Mode of Wireless Video Sensor Network

The Data management mode of wireless video sensor network acquired by urban road traffic flow information is mainly supported by protocol design. The system protocol design adopts the frame of Fig. 2, and the sensor nodes use improved protocol based

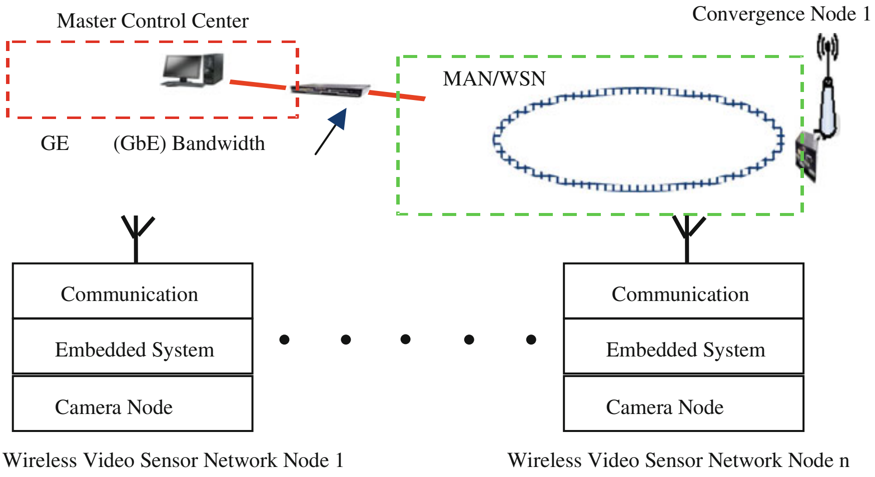


Fig. 1 The architecture of wireless video sensor network acquired by traffic flow information

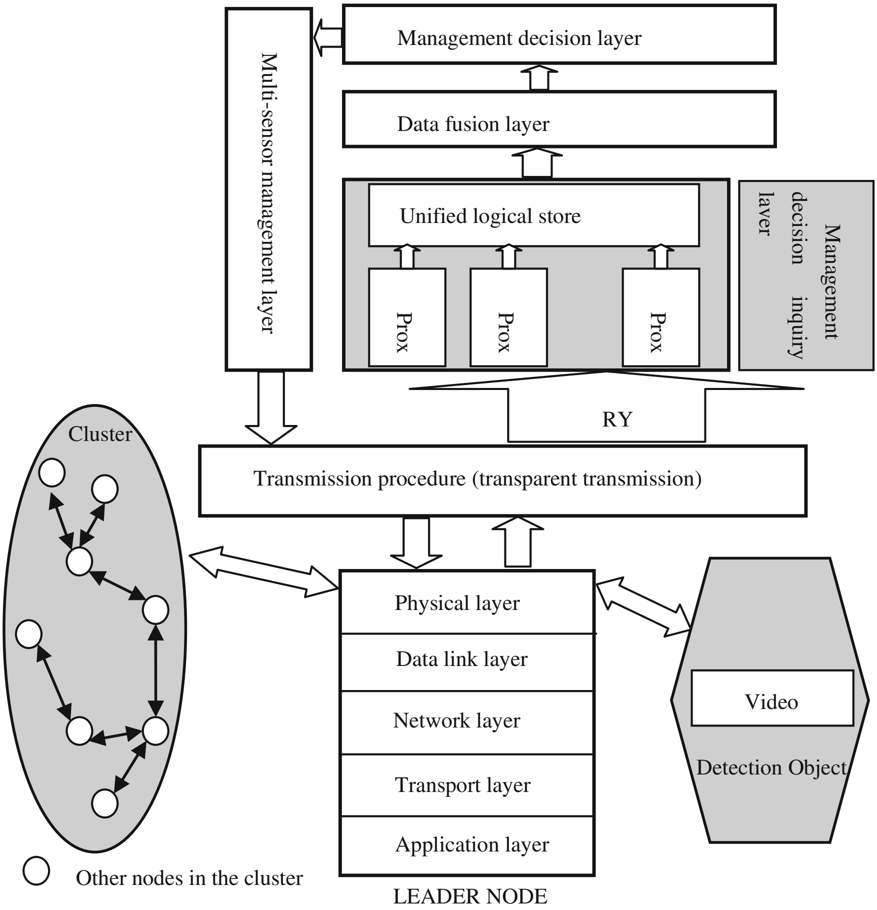


Fig. 2 Protocol design frame of the system

on cluster. The cluster LEADER includes application layer, transport layer, network layer, data link layer and physical layer. The cluster LEADER node comprises self-organizing network with other nodes in the cluster, meanwhile, the communication with communication infrastructure is in a transparent status [8–10].

Above the transparent procedure, they are mainly responsible for the data inquiry management, fusion, decision making, and management for each parameter of the sensor nodes in the detection area according to the decision making. Data inquiry and management layer is divided into normalization logical storage and each inquiry agent according to specific division of labor, and they transfer users' inquiry commands to nodes by normalization, where they obtain needed information. After that, information got from the data fusion layer is fused in accordance with decision making, then, the results are delivered to management decision making layer.

Multi-sensor management layer configures each parameter of the sensors in each area judging by the results and the requirements of the decision making in order to satisfy the next detection demands. Meanwhile, the commands are delivered to LEADER node by transmission network, and executed in the cluster based on specific situations [10, 11].

4 Streaming Media Transmission and Access Applications in Wireless Video Sensor Network

The design and realization of the hardware system of the traffic flow video detection node based on wireless sensor network and wireless streaming media technology adopt the self-design platform combing high-end ARM embedded system with programmable logic device FPGA (field programmable gate array). Using 2.4 G wireless communication, high-end ARM embedded system designs three functions:

A. Control function: it controls video image data acquisition unit, which includes initializing set up of acquisition, the choice and adjustment of the characteristic parameters in the acquisition procedure, the streaming rates in the acquisition procedure and so on.

B. Data exchange function: After successfully gathering the image data of the video image data acquisition unit, CPU provides data exchange function. The high speed read characterized by DMA (Direct Memory Access) reduces the participation of the CPU to the lowest level, which makes the read to the image data no longer need to occupy the resources of the CPU.

C. Data processing function: it is used to non-DSP related IP network processing, which provides IP communication related processing capacity and function. Its performance is to effectively and rapidly process network related affairs, which provides IP exchange conditions for digital camera hardware system.

The programmable logic device FPGA platform can achieve the function, that is, the quick channel of the high-speed image acquisition system unit and CPU need to be built. At present, there is no IC chip of no glue logic interface technique at home and abroad, the interface based on FPGA can be effectively set up, while the high-speed image acquisition system unit and CPU can be established. Such data channel concludes two main data flow: first, it is the data flow after image acquisition with black and white, color information, related timing signal and so on. Second, it is working status information of CPU control data and image acquisition sub-system, which reduces unnecessary soft processing work of CPU, and makes system work more effectively and fast. Third, FPGA platform can process data and connect with ARM effectively.

The design and realization of wireless sensor network streaming media transmission and access adopts three-level system structure. Where, multi-capacity IP broadband multichannel video network in the core of wireless sensor network and wireless transmission technology are researched and realized, while the networking is convenient, monitoring and management are sharing the same wireless broadband network. The three-level structure is as following: no wired work of intelligent compound-eye digital camera system and the fusion of wireless detection sensor are the first level, the design and realization of wireless network bridge is the second level, which achieves wireless connection of multimedia data flow, interconnection with one or multi-block roads, where the data can be transferred, command instructions are unobstructed. The IP access technology is the third level, which is compatible with wired IP network and optical fiber, and it supports multi-layer management integration and achieves easily extended intelligent wireless video system, meanwhile, it supports one-level to multi-level control centre which can detect multi-channel multi video flow, and transfer multi-channel video flow of each channel in order to realize grading and separation of powers, and run safely and steadily in a large area to meet the requirement of easily maintain [3, 12, 13].

In the view of wireless communication, the design used in this paper adopts up-to-date international standard IEEE802.11n and related compatible standard (Backward Compatible IEEE802.11b/g) to realize the second and the third level wireless communication, meanwhile, according to the domestic actual condition, streaming media transportation uses IP protocol (TCP/IP, UDP/RTP/RTCP, etc.) by improving and choosing WiMAX standard. In the design of the base station, the open system structure is used as main structure, and the compatibility with the current system is emphasized to achieve smooth transition and system improvement. The design and the realization of the system adopt the advanced technology that is researched by ourselves, which provides advantages for preliminary design work, and effective platform for sustainable development of media and long term design work.

5 Conclusion

With the development of micro electronics, computer vision and image processing technology, as a traffic flow information acquisition method with development potential, video detection can acquire traffic flow information such as traffic flow, vehicle length classification, time occupied rate, space occupied rate, headstock distance, comparing to other detection methods such as induction coils, ultrasound, microwave, infrared and laser.

This paper introduces recently new developing wireless multimedia sensor network technology, carries out key technology researches of video detector with new large field of view, illumination adaptive well, real-time, high detection rate and easy to install and deploy. Meanwhile, it proposes a design mode of streaming media transmission and access application in wireless video sensor network, which provides a guide and reference for solving traffic flow information acquisition.

References

1. Luo W-S, Zhai Y-P, Lu Q (2008) Study on wireless multimedia sensor networks. J Electron Inf Technol V30(6):1511–1516
2. Bell MGH (1992) Future direction in traffic signal control. Transp Res A 19A(5–6):369–373
3. Zhong S, Zong C (2010) The application of streaming video monitoring in highway supervision. Transp Inf Ind 5(7):38–44
4. Ding X, Xu L (2011) Stereo depth estimation under different camera calibration and alignment errors. Appl Opt 50(10):1289–1301
5. Wang H (2010) Adaptive down-sampling stereo video coding based on background complexity. J Inf Comput Sci 7(10):2090–2100
6. Xiaofeng D, Lizhong X (2011) Robust visual object tracking using covariance features in Quasi-Monte Carlo filter. Intell Autom Soft Comput 17(5):571
7. Lizhong X et al. (2011) Trust region based sequential Quasi-Monte Carlo filter. Acta Electronica Sinica 39(3):24–30
8. Islam ABMAA, Hyder CS, Kabir H, Naznin M (2010) Finding the optimal percentage of cluster heads from a new and complete mathematical model on LEACH. Wirel Sens Netw 2(2):129–140
9. Pirzada AA, Portmann M, Indulska J (2008) Performance analysis of multi-radio AODV in hybrid wireless mesh networks. Comput Commun 31(5):885–895
10. Wang H, Lizhong X (2007) Route protocol of wireless sensor networks based on dynamic setting cluster. In: Proceedings of 2007 IEEE international symposium on industrial electronics, Vigo, Spain, June 4–7
11. Nazir B, Hasbullah H (2010) Energy balanced clustering in wireless sensor network. In: International symposium on information technology (ITSim), Kuala Lumpur
12. Tan G (2011) Optimizing HARQ with rateless codes for wireless multicast under strict delay-bandwidth constraints. Adv Inf Sci Serv Sci 3(7):347–356
13. Li Y et al. (2011) A new method for improving the small size spacing four-element MIMO system channel capacity and its stability 1(6):1–6

PSO-PAC: An Intelligent Clustering Mechanism in Ad Hoc Network

S. Thirumurugan and E. George Dharma Prakash Raj

Abstract The impact of wireless network has been growing day by day. The communication without any infrastructure has been the phenomenon for the present and future has been felt and realized. In line with that, ad hoc network with clustering mechanism proves an improved result. The existing algorithm considers either single parameter or multiple parameters to form clusters using IPv4 nodes. This work proposes PSO-PAC as an optimized clustering mechanism with the help of swarm intelligence to devise the clusters by considering the vital parameters. This study takes crucial parameters to form clusters dynamically using IPv6 configured nodes. This paper work has been supported by the implementation using OMNET++ as a simulator.

1 Introduction

The wireless network places an important role in present situation. This network with an improved performance has been indispensable since the scalability will limit the functionality. In order to put the efficiency value of the ad hoc network to the high value multiple parameters are being considered. It is obviously realized that the distance parameter alone couldn't decide the efficiency of clustering mechanism. Since the nodes which are in wireless network are facing energy drain as a problem with respect to time. Further, the cluster head needs maximum energy among the nodes in a cluster to act as a transceiver. The existing distance based algorithms are

S. Thirumurugan (✉)
Department of Computer Applications, JJ College of Engineering & Technology,
Tiruchirapalli, India
e-mail: s_thiru_gan@rediffmail.com

E. George Dharma Prakash Raj
Department of Computer Science and Engineering, Bharathidasan University,
Tiruchirapalli, India
e-mail: georgeprakashraj@yahoo.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_6,

showing deficiency in giving an optimum solution. Thus, this work proposes multiple parameter based algorithm PSO-PAC (Particle Swarm Optimization of Partitioning Around Clusterhead). The swarm intelligence has been the base for the formulation of this efficient clustering mechanism. The behavior of the crow has been considered to devise the procedure of this technique.

This paper has been organized as follows. Section 1 deals with introduction. Section 2 says about the related works. Section 3 tells about the Ex-PAC mechanism. Section 4 gives out the proposed PSO-PAC procedure. Section 5 deals with the experimental results. Section 6 puts down the future works. Section 7 ends up with the concluding remarks.

2 Related Work

The purpose of clustering has been realized when the protocols like AODV [1] has been integrated with the clustering mechanism. Since this addition to the existing protocols produces an improved result.

The application of the clustering mechanism [2] with AODV as a routing protocol in the real world scenario has been unavoidable. This shows that clustering technique makes the network to be suitable for various real world applications. The clustering mechanism PAC [3] over the k-means approach tells the purpose of parameters in cluster formation. These parameters decide the efficiency level of clusters. This study also confirmed that k-means takes more time when the number of nodes are high in count. This work lacks in implementation and also the sample set of nodes are small in size.

The PAC procedure worked well for less number of nodes. But for more number of nodes this has left many nodes as non clustered nodes. The Ex-PAC [4] came out as an extension to PAC which takes entire nodes and produces the maximum possible clusters. The cluster formation process eventually improved in Ex-PAC procedure. This approach concludes that Ex-PAC has shown significant improvement over k-means in terms of computational speed.

The cluster formation algorithms devised so far lacks in obtaining an efficient cluster in terms of time, maintenance and selecting proper re-clustering phenomenon. The IPv4 address no longer will serve the world has been realized. The IPv6 address has got the focus in the present scenario. The configuration of address can be done in two ways. The configuration using DHCPv6 server and stateless autoconfiguration [5]. This configured address can be local link address or global address. This will be decided based on the application requirement. But there must be awareness on the limitations of using the address range has been vital.

The stateless autoconfiguration of IPv6 nodes should have duplicate address detection mechanism. Since when the nodes move across the clusters there is a chance of address duplication. To eliminate this passive duplicate address detection mechanism [6] has been introduced. This method shows better results than passive autoconfiguration for mobile ad hoc networks.

The purpose of autoconfigured address has been recognized while the ad hoc network has come to reality. The limitation of IPv4 and the need of IPv6 [7, 8] has been understood clearly. Having understood the difference between IPv4 and IPv6 the scenarios will demand the specific way of addressing the nodes. These works are dealing with the autoconfiguration or manual configuration of IPv4 and IPv6 nodes. The clustering as a mechanism comes for this IPv4 or IPv6 autoconfigured node to make the routing simple and also confirms the efficient utilization of bandwidth and resources.

3 Ex-PAC

The PAC creates the clusters based on Manhattan distance. The Manhattan distance saves time in computing the distance between pair of nodes. The results achieved are not adequate to find out the appropriate cluster in the ad hoc scenario network. This has been further enhanced through Ex-PAC algorithm which has been laid on top of PAC. The experimental results show that Ex-PAC has given better results than K-means [9] algorithm in forming clusters. The formula (1) puts down the calculation of Manhattan distance.

$$\text{Manhattan Distance} = \sum_{i=1}^{n} |xi - yi| \tag{1}$$

Ex-PAC algorithm

```
(1) Assume Ni = temporary Cluster head.
(2) Compute Manhattan distance between pair of nodes.
(3) if (MD < range)
        Begin
            Add (Ni, Ci)
            Count = Count + 1
        End
(4) Repeat the steps 1 through 3 till all the nodes in the cluster are examined.
(5) Select the First Cluster which has maximum count value.
(6) Select the non cluster nodes.
(7) Choose the cluster which includes non cluster nodes.
(8) Select cluster returned in step 7 as Second Cluster.
(9) Repeat the steps 6, 7 and 8 until there is no change in Cluster formation.
```

The Ex-PAC procedure has produced remarkable improvement over the PAC procedure. But still lacks in ensuring the perfectness of the clusters. Thus, the validation parameters are needed to be considered to check the integrity of the clusters.

4 PSO-PAC

The crow behaves distinctly when it finds eatables. It alerts others after seeing the eatables through the acoustic signal. Thus, a group of crows will devise the cluster. This formation is based on more than one parameter. Those parameters are distance and energy of the crows in communication process. This multi parameter specifies that the cluster formation with the help of swarm intelligence makes the clustering process in a highly efficient way. This behaviour of the crow can be put in making dynamic clusters.

The Fig. 1 shows the network model consist of two clusters where each cluster has separate crow head node. The gateway acts as mediator between two clusters. This ensures inter-cluster communication to happen. In this model the destination is far away from the source. This network setup will be retained till the communication between the source and destination gets over. After the data transfer gets over then network has to be reformed with new energy source and destination. In this way the cluster formation happens completely dynamic in nature which suits very well with the behaviour of the crow.

4.1 Cluster Formation by Crow Behavior

```
(1) Input the clusters formed in Ex-PAC procedure.
(2) j = j + 1
(3) Cluster = Cj
(4) i = i + 1
(5) E = Get_energy (Ni, Cluster)
(6) If (E < EnergyThreshold)
        Newcluster (Kj, Ni)
(7) Repeat the steps 4, 5 & 6 until all the nodes belong to the cluster Cj gets examined.
(8) Repeat the steps 2 through 7 until all the clusters of network are examined.
```

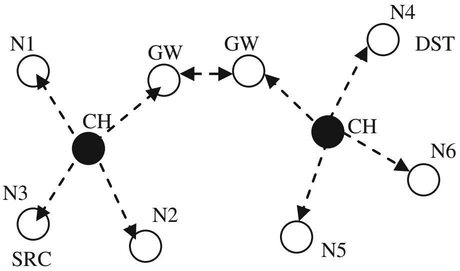


Fig. 1 Network model:two clusters

4.2 IPV6 Configuration of Clusters

The cluster formation takes the output of Ex-PAC procedure and recreates the cluster dynamically based on the energy level of nodes. The Fig. 2 shows the IPv6 nodes to form the clusters dynamically in a stateful way.

IPv6 Configuration Procedure

```
(1) Input the clusters from crow behaviour procedure.
(2) Install the DHCPv6 server process in Cluster head node.
(3) Cluster head sends hello message to all the members.
(4) For each Cluster Ci
     Begin
       For each member node Nj
         Begin
           Node Nj sends reply to the Cluster head.
           Cluster head sends an IPv6 address to Nj.
         End
     End
(5) Repeat the step 4 for j = 0 ... Number of Nodes in cluster Ci and for i = 0 to
    Number of clusters.
```

The IPv6 configuration procedure configures the nodes belong to the clusters under stateful way using DHCPv6 server. In this method the cluster head sends hello message to all the members at the initial stage. The member nodes send the reply to the cluster head. This reply informs the member belong to the cluster where the cluster head exists. The cluster head now sends IPv6 address to the member node as a part of configuration.

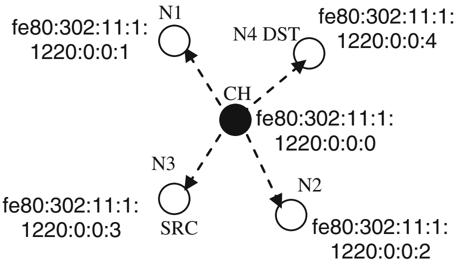


Fig. 2 IPv6 configured network model

Table 1 Simulation parameters

Parameter	Values
N (number of nodes)	25, 50
Space (area)	100 × 100
Tr (transmission range)	20 m
Ideal nodes percentage	90 %
Simulation time	5 s
Threshold(energy)	500

5 Experimental Results

This PSO-PAC algorithm has been implemented in OMNET++ and the results are tabulated. This work has been carried out with the system configuration of 64 bit AMD processor, 2 GB RAM and windows XP as an operation system. This simulation has been done for 25 nodes and 50 nodes.

The Table 1 shows the simulation parameters considered while simulating the study. The cluster heads could be fixed and need to be changed in accordance with the time T. The energy level graph Fig. 3 at a specific time T1 identifies the node8 has been an eligible node to be considered as cluster head. After time period T1 some other node will be chosen as cluster head at time period T2. This graph clearly reveals the cluster heads selection in the order. The Fig. 4 shows the nodes and their respective energy level belongs to cluster C2. This ushers that the node21 will be chosen as cluster head since it has high energy level at time T1.

The Fig. 5 ushers the energy based results for the sample size of 50 nodes. This shows that node8 takes the role of cluster head.

The Fig. 6 ushers the energy based results for the cluster2 where the sample size has been 50 nodes. This identifies node39 as the cluster head.

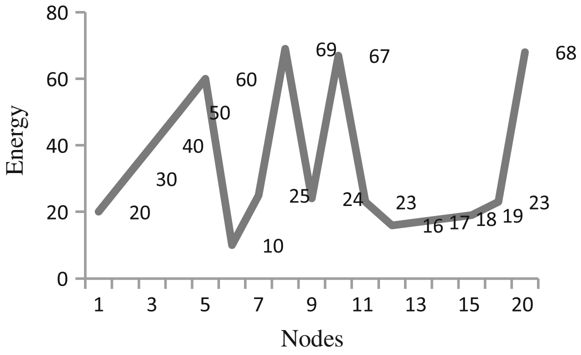


Fig. 3 Energy graph for 25 nodes: cluster C1

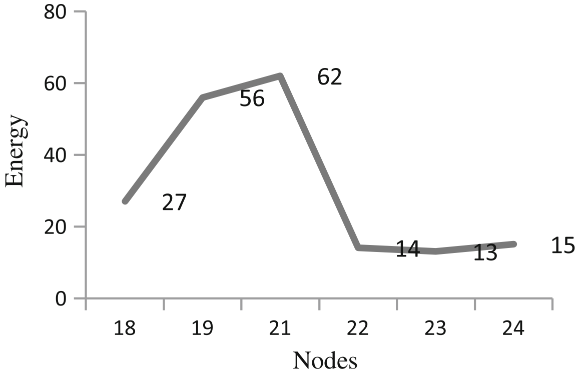


Fig. 4 Energy graph for 25 nodes: cluster C2

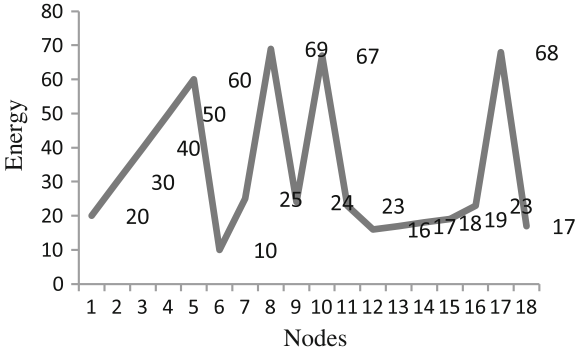


Fig. 5 Energy results for 50 nodes: cluster C1

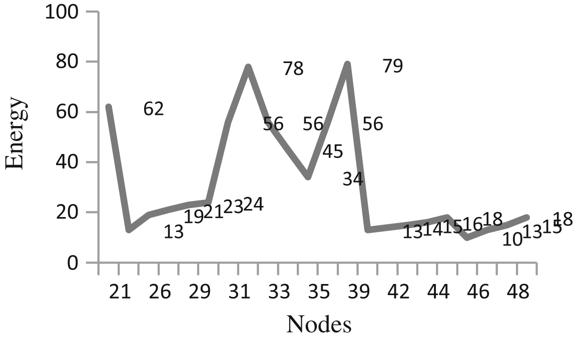


Fig. 6 Energy results for 50 nodes: cluster C2

6 Future Directions

This work considers the stateful configuration of IPv6 nodes. The future work should also consider the stateless auto configuration of IPv6 nodes to form the dynamic clusters in ad hoc networks. The implementation of mobility of nodes or cluster head should also be considered.

7 Conclusion

This study considers the dynamic cluster formation. The intelligence of crow has been simulated on formation of the dynamic cluster. These clusters tentatively exist till the data transfer gets over which has been similar to crow finds eatable, devising cluster of crows and eats off the food. This study gives an understanding of the crow's behavior to be applied on formation of clusters in ad hoc network. This work has been simulated using OMNET++ to substantiate the results.

References

1. Thirumurugan S (2010) Direct sequenced C-IAODV routing protocol. Int J Comput Sci Technol 1(2):108–113
2. Thirumurugan S (2011) C-AODV: routing protocol for Tunnel's network. Int J Comput Sci Technol 2(1):113–116
3. Thirumurugan S, George Dharma Prakash Raj E (2011) PAC—a novel approach for clustering mechanism in ad hoc network. In: ICSCCN'11, pp 593–598, 21–22 July 2011
4. Thirumurugan S, George Dharma Prakash Raj E (2012) Ex-PAC: an improved clustering technique in ad hoc network. In: RACSS'12, pp 195–199, 25–27 Apr 2012
5. Qing L (2010) An IP address auto-configuration scheme for MANET with global connectivity. In: ICIME'10, pp 244–247, 25–26 Mar 2010
6. Sivakumar M, Jickson CJ, Parvathi RMS (2011) Passive duplicate address detection in DYMO routing protocol for MANETS. In: ICVCI'11, pp 9–14, 7–9 Apr 2011
7. He Z, He Y (2010) Study on key technologies of MANET. In: ISME'10, pp 112–115, 7–8 Aug 2010
8. Herberg U, Clausen T (2010) Yet another autoconf proposal (YAAP) for mobile ad hoc NETwork. In: MASN'10, pp 20–26, 20–22 Dec 2010
9. Wu J, Yu W (2009) Optimization and improvement based on K-means cluster algorithm. In: Second international symposium on knowledge acquisition and modeling, pp 335–339, 30 Nov–01 Dec 2009

SEMSuS: Semantic Middleware for Dynamic Service-Oriented Sensor Network

V. Sangeetha and L. Jagajeevan Rao

Abstract Sensor network appears to be the technology of the 21st century. Middleware for wireless sensor networks (WSN) acts as the interface point between wireless sensor nodes and higher application layers. A middleware can introduce and enhance access to the underlying networks in intelligent ways. In this paper we present a middleware for wireless sensor networks that uses a set of technical statements, such as patterns and styles, in order to achieve flexibility, compatibility, autonomy and adoption. The middleware exposes the functionality of the network as semantic web services, so that applications can access its functionality through web services. Sensor web combines sensors and sensor networks with a Service-Oriented Architecture (SOA). The Service Oriented Architecture allows us to discover, describe and invoke services from a heterogeneous software platform. Dispatcher and Interceptors are used inside the network. We propose a scenario in which two services are exposed a semantic web services and designed to run in a constrained environment and they are exchanged in accordance with capabilities of the network.

Keywords Wireless sensor network · Middleware · Service oriented architecture · Semantic web services

1 Introduction

The main purpose of Middleware for sensor networks is to support the development, maintenance, deployment and execution of sensing-based applications [A]. The emerging technologies and concepts to increase the flexibility of software solutions

V. Sangeetha (✉)
KL University,Vaddeswaram, Guntur, India
e-mail: sangeethalokanadam@hotmail.com

L. J. Rao
School of Computing, KL University, Vaddeswaram, Guntur, India
e-mail: jeevan@kluniversity.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_7,

in various contexts have emerged in computing. It will potentially add millions of new data sources to the web. In distributed systems, mechanisms to access, reflect on and change their own interpretation is a reality [1]. At run time the systems are reflexive, in which reflection is related to the ability to inspect and adapt the system at dynamic process [2].

This paper aims to show the design, implementation, and execution of the SEMSuS (Semantic Middleware for Dynamic-Service-Oriented Sensor Network). It allows run-time adaptation of the functionalities that run on its platform. Adaptation as considered in this work results from the adoption of architectural styles, such as Service-Oriented-Architecture (SOA) and design patterns, such as the Dispatcher [3] and Interceptor [4].

Reflection can be applied in different types of systems, such as in wireless sensor networks (WSN), but requires considerations concerning the intrinsic characteristics of these networks. A WSN consists of several small nodes with restrictions on computing power, the bandwidth of wireless communication and the amount of energy available [5]. Therefore the gradual increase in computational cost, as a result of the addition of flexible mechanisms, must be minimized in order to extend the lifetime of the network. Sink node is one type of the network gateway, so that the communication node sources to any external system network should be through the sink.

Another feature of WSNs is that they are closely related to the environment in which they are deployed. The sensor nodes are scattered in an environment of observation for the extraction, processing and sending data to requesting information outside the network through one or more sink nodes.

When using SOA, the WSN plays the role of service provider, while the application is its client. Associated with the need for adaptation and the characteristics of WSNs, is the solution for middleware platforms capable of performing various types of applications. In order to maximize the automation of activities related to services, they are implemented as Semantic Web Services (SWS). Then a network service can be found by matching semantic. After the choice semantics is invoked at runtime, a specific implementation of the dynamic service among the network.

In Sect. 2, we look at the concepts and technologies needed for future enhancement of the work, such as SOA, WS, SWS and Ontologies. Section 3 gives the clear explanation of middleware SEMSuS. In this section are first analyzed the structural components and then the interactions between these components.

In Sect. 4 we show the use of middleware in order to evaluate the implementation and finally in Sect. 5 presents the conclusion and the future work.

2 State-of-Art

Similarly to that presented in [1], a WSN can be considered from the perspective of a provider of services that meets the needs of applications running on its platform. The adoption of services makes transparent its implementation and standardized access features, allowing the addition of flexibility and dynamism to the solutions, becaused different services can be selected and exchanged it with no change in the form of client interaction. Adding semantic Web technologies to the services description promotes greater degree of automation by making improvements in the mechanisms of selection, composition, invocation and monitoring of services.

2.1 Dynamic Service-Oriented Architecture and Web Services

The identification of common data operations and transformations on sensor data has to introduced the sensor web paradigm. Sensor Web combines Sensors and Sensor networks with a Service Oriented Architecture (SOA). A SOA allows us to discover, describe and invoke services from a heterogenous platform using XML and SOAP standards. The components are available as independent services that are accessed in a standardized manner [6]. There are three main actions that are performed in SOA: publish find and bind/invocation. To support them there are three main components the client, the provider and the registry. Figure 1 represents the SOA with the components (in blue), and their interactions (yellow) in the sequence.

The basic interactions or activities are: (1) Publish—the provider makes the publication of service they can provide in a place known by the clients (the record); (2) Find—The client requests information about the services published in the registry that can meet their needs; (3) Bind/Invocation—possession with the address of the service that meets its needs, the client interacts the service provider to deliver it.

Web Services (WS) are the most popular realization of SOA [6]. The invocation of the WS is made through XML messages [7] that follow the SOAP standard [8]. The operations contained in a Web Service, as well as its inputs and outputs parameters are described in WSDL [9]. The service's WSDL is published in a registry called Universal Description, Discovery and Integration (UDDI), which is consulted by clients seeking services. This use of a set of standards makes interoperable the WS.

2.2 Ontologies and Semantic Web Services

An Ontology is a formal and explicit specification of a shared conceptualization [10]. Where "formal means readable by computers, explicit specification with regard to concepts, properties, relations, functions, constraints, axioms explicitly defined,

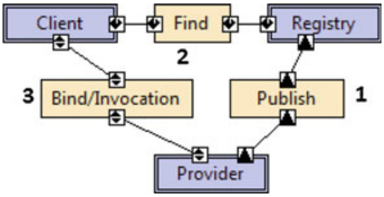


Fig. 1 Service oriented architecture

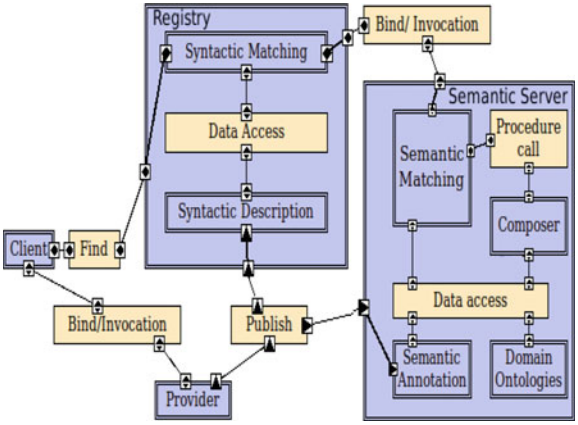


Fig. 2 Reference architecture for SWS

shared knowledge that is consensual, and conceptualization concerns an abstract model of some real-world phenomenon" [11]. So, an ontology defines concepts and relations between them in a field of knowledge.

In the Fig. 2, the more internal components shows the division from the logical viewpoint, while the external shows the deployment viewpoint. In this architecture there are two matching in different locations: one syntactic, in the Registry, and other semantic, in the Semantic Server. The Registry performs the matching comparing, syntactically, the client request with the relevant service publication, while the Semantic Server performs the semantic matching from the semantic descriptions of both services in the requests. To carry out the semantic descriptions it is necessary to use domain Ontologies for semantic annotation.

The SWS is designed by the addition of semantic technologies to the patterns of WS, promoting the enrichment of the services description and behavior. Figure 2 shows a reference architecture that expresses the operation of the SWS. The mechanisms that allow the services provided by a WSN to be accessed through Web Services provide interoperability through a standards-based framework for exchanging information. Such patterns for WS are designed to represent the service interface, how they are distributed and how they are raised. However there is a limitation in expressing what services do. The automatic integration is compromised because it is based solely on the syntax and semantics rather than a description of the service [6].

3 Related Work

To design a successful dynamic Service-Oriented Architecture for sensor network, each classification focuses on different aspects and different purposes. Some of the main middle wares for WSN are Agila [12], MATE [13], TinyDB [14] and Tiny Lime

[15] among others. More specifically there are approaches that make use of SOA, an can be found in [1]. Semantic Web technologies relate to other works as in [16, 17].

The SOA-based approach proposed in SEMSuS intended to allow flexibility in adding new services in a manner similar to that implemented in [1], which shows a set of services, among which service configuration and adaptation of the network that allows activation of adaptation policies when any change occurs in the environment. Differences in the middleware proposed work in relation to SEMSuS is that it has as the execution platform the Java Platform Micro Edition (Java ME) [18], which are less stringent as that based on Mica motes. So there is a greater capacity to perform the middleware in the nodes. Another difference in the middleware proposed in [1] is that its implementation is only in the nodes, while in SEMSuS middleware is distributed by a number of devices that transcend the limits of WSN, such as a web server, allowing load balancing middleware so that they can be used Mica nodes in the network. Another difference is that in [1] the middleware does not use semantic web technologies in their mechanisms.

A work that makes use of semantic services is [16] which use ontologies to represent data from sensor nodes, as well as its services, and allows the descriptions of the services have semantic annotations. The paper presents a three-layer model, in which reside the Platform, which is related to the hardware and operating system. The software layer allows making transparent the heterogeneity of the platform, and the semantic layer makes transparent the semantic heterogeneity of the Software. An interesting fact in [14] is a proposal to make use of Ontologies that describe the data and services in WSN. That is, each node can represent its skills and its data through the domain ontology, which are presented and analyzed in the work. The use of Ontologies allows nodes of different platforms, and using different representations of data, can communicate with each other made the network more flexible.

The difference between the work in [16] and SEMSuS lies in the fact that the mechanisms that use Ontologies to describe data and services are performed in the environment of the WSN. In SEMSuS the publication and use of Ontologies comes at a higher level of architecture because of the limitation of sensors that do not allow the use and processing of files such as OWL [19]. Another factor is that the proposal in [16] is performed to describe the capabilities of each node of a WSN, while in SEMSuS is made to describe functionalities of a WSN.

Another work is [17] which makes use of semantic web technologies to build a semantic middleware for autonomic WSN. The system that allows autonomy for network uses a set of rules of inference on Ontologies and logic-based fuzzy. Data networks are mapped to domain Ontologies and rules act on the Ontologies, which provide a high level of abstraction to the applications. Locus inferences also run on sensor nodes in order to provide necessary data for a set of services that implement the autonomous capabilities. The difference in [17] and SEMSuS is that the services that implement the autonomy using data provided by inference engines, which does not occur in SEMSuS. Another difference is that the SEMSuS uses SOA for providing means of access to the WSN via the web.

4 The SEMSuS

The SEMSuS is a middleware for WSN that aims to present sensor networks as providers of semantic services. The adoption of SWS allows client applications and networks to exchange information that can be semantically interpreted by both. Because the use of late binding technique promoted by the Manager and Dispatcher, the SEMSuS choose the implementation of the service according to the network capacity, i.e. if the network is with fewer resources is chosen implementation that consumes less resources at the expense of guarantees, for example, the response time.

In SEMSuS, services are implemented inside the nodes by means of language nesC [20] on the Tiny OS operating system [21], which were especially developed for constrained environments. The selection of service implementation in WSN is held in the main network component that implements the standard Interceptor, which is called the Manager.

4.1 Architecture of Middleware

The architecture of middleware is designed by specifying the provider components in the lowest level that form the provider. One of these components is the Web Server and the other is the WSN. The Provider architecture is depicted by the Fig. 3.

The Web Server component is a server that has the functions to receive requests from client applications and publish the capacity of the existing networks in the Registry. While the server does not have resource constraints, the other component that binds it (WSN) possess constraints, since it is performed by Mica nodes. Therefore, the semantics that connects them is an adapter between two different contexts.

The gateway implements the main features of the Web Server to the system. As its name suggests, it is the component for which the provider performs the interaction with the others external components. Such interactions are reflected by the publication of services in the Registry and Bind/Invocation of the client.

In WSN there are two important components: the Sink and Source Node. The Sink is the drain of the WSN data, so that any network communication with the external environment occurs through it. Usually the Sink node has more resources as compared to the Source Nodes.

In the Sink there are two logical components: the Manager and Dispatcher. The first one has the capability to manage and choose the identifiers of each service implemented by sensor nodes. Management concerns to placing an appropriate value that identifies a service so that the Dispatcher can then call run-time algorithm that really implements the service.

In sources nodes components is the implementation of services. All services follow a common interface. Then the addition of new services becomes as easy as possible. The added service must implement the interface system and have an identifier associated with it. Then the service can be invoked when the Dispatcher is

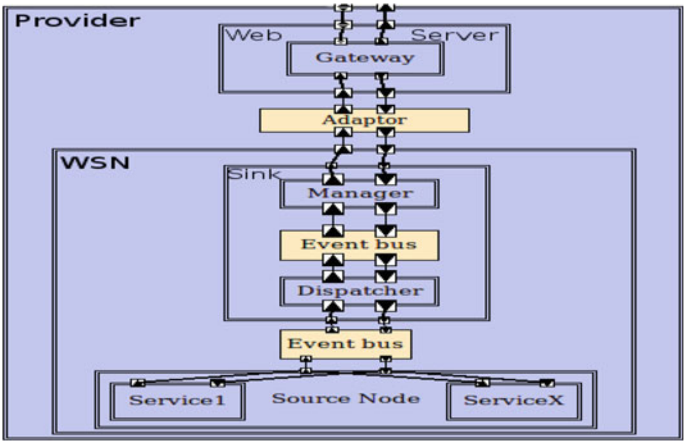


Fig. 3 Reference architecture of the provider

required. When removing a service is required to be removed its identifier, so that so that the service is no longer invoked.

The opening of the middleware architecture allows, for example, a service that monitors the state of the WSN, and to report to the Manager the current network conditions, can be added so it can choose the best implementation for the requested service. So the Manager acts as an interceptor that can interrupt the control flow allowing it to run other code.

4.2 Interaction Between the Main Components of Middleware

Figure 4 shows the interaction between the main components of SEMSuS. The process begins with the application seeking the service that can meet its requirements. This search can be performed either in a syntactic or semantic. The two options are provided through the implementation of the API (Application Programming Interface) allowing, thus, the flexibility of choice for application development. If the service is found, their identification is returned for later invocation. By holding the address of the service provider and knowing its interface, the application relies on passing the necessary data and receiving results of its implementation. In implementing the system, the Registry is performed by the jUDDI tool and the Semantic Server by Matchmaker. The jUDDI implements the syntactic matching and the Matchmaker implements the semantic matching.

Another important activity is the invocation of service for the client application. The interactions between the components involved in this process are presented in Fig. 5. It can be possible to observe that the process begins with the invocation of the application by the service provider. The first software component in the provider

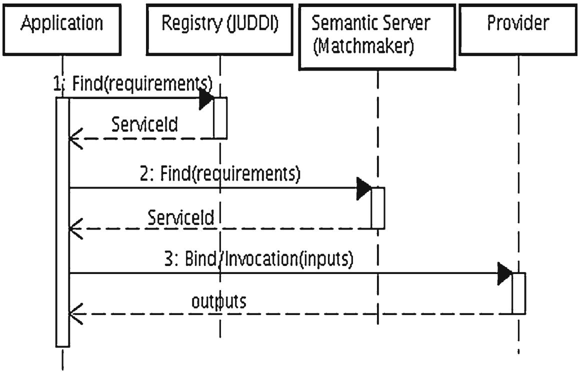
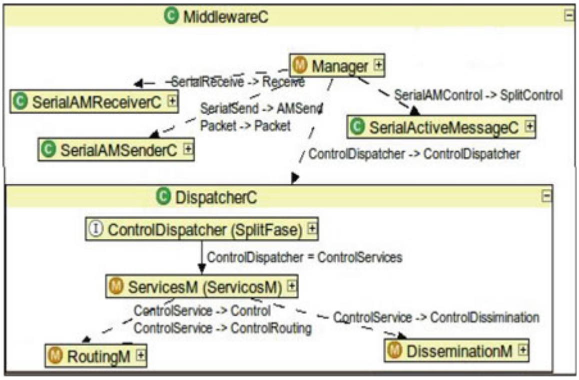


Fig. 4 Interaction between main components of the system

Fig. 5 Interaction between invocation of services


MiddlewareC
Manager
SerialReceive -> Receive
SerialAMReceiverC
SerialSend -> AMSend
Packet -> Packet
SerialAMControl -> SplitControl
SerialActiveMessageC
SerialAMSenderC
ControlDispatcher -> ControlDispatcher
DispatcherC
ControlDispatcher (SplitFase)
ControlDispatcher = ControlServices
ServicesM (ServicosM)
ControlService -> Control
ControlService -> ControlRouting
ControlService -> ControlDissimination
RoutingM
DisseminationM

that will receive the requests is the Gateway, which is a web server Apache Tomcat. In the Gateway, the requests are transformed into messages to be sent to the WSN.

In WSN component, the Manager component receives the message from the Gateway then sends the request with the identifier (id) code that implements the service requested by the Client. The choice of service (id) depends on the current capabilities of the network. The results of the implementation of the service are returned in the opposite direction until reach the Client. Thus, the process allows the invocation of the code in a WSN as SWS.

5 Implementation and Usage Issues

To validate the SEMSuS, was implemented a scenario in which two network services are exposed as SWS. Both services have the function of communication in a WSN, but with different characteristics. One performs a communication by means of routing and another by food of messages. Figure 6 shows the main components implemented in nesC language.

The Manager is the first component in the WSN to receive requests and it is in the sink. The Manager uses other components, such as SerialActiveMessage, to deal with the communication via the serial port, which connects the Sink to the Web Server. Because all services implement the same interface, the ControlDispacher, they can be switched without major problems.

The routing service is the Tymo, which is based on the Dymo (Dynamic MANET On-demand) for TinyOs. The Tymo inherits the main features of Dymo, such as point-to-point communication and find routes to hosts in the network. Therefore, messages for recognition of the paths are disseminated in network only when a node needs to communicate with another. The memory savings is due to small amount of data stored by the communication and energy "on-demand". Furthermore, the Dissemination service is based on the dissemination protocol to send data to all nodes in the network constituents. It works by flooding the network with data.

The routing has the advantage of saving more network resources, as it decreases in the average number of messages that are exchanged, saving bandwidth connections and energy of the nodes, when compared with the dissemination.

Moreover, the dissemination to a lower response time, since it is not necessary to exchange messages to the knowledge of the route. Thus, if the network has sufficient resources, the dissemination should be chosen to implement the service which governs the form of network communication.

In the interaction between the components of SOA, the client application uses the Matchmaker Client to conduct the semantic search, and OWL-VM to invoke the SWS that are published in OWL-S, in the registry.

6 Conclusion and Future Work

Software solutions found for distributed systems are often complex and heavy. This fact is due to the generality of Middleware transparency in dealing with the distribution to the most varied requirements of different applications. The adoption of open systems, flexible and adaptable is proving useful in the treatment of this fact.

In middleware proposed in this paper, the SEMSuS uses patterns and architectural styles consolidated to achieve openness, flexibility and adaptability. Concern to the SOA style is used as a form of interaction between the WSN and applications, specifically its implementation is done through SWS. The application is the client while WSN is the service provider. The adoption of SWS offers flexibility and dynamism. Not just the interaction between network and application are the mechanisms that promote flexibility, but also within the network. This fact is achieved by using the Manager and Dispatcher components, which implement, respectively, the patterns Interceptors and Dispatcher. As a result there is possibility of changing the implementation at runtime of the service provided by the network according to the capabilities of the network.

Within the architecture, the Manager component has key role by selecting the service implementation. This choice is deterministic. Different approaches of choice can be made to support the Manager, including some approaches a more flexible programming, such as those based on logical or on Ontologies and inference engines. Implementations of such approaches can be easily added in the SEMSuS as services, if the implementations satisfy the interface of Dispatcher.

References

1. Delicato FC (2005) Middleware-based services for wireless seNsor netwok. Doctoral thesis of the Federal University of Rio de Janeiro, Brazil, Rio de Janeiro, p 53, June 2005
2. Capra L, Emmerich W (2001) Mascolo C reflective middleware solutions for context-aware applications. Proceedings of the reflection 2001, Lecture notes in computer science 2192, Springer Verlag, Japan, pp 126–133
3. Gay D, Levis P, Culler D (2007) Software design pa tterns for TinyOS. ACM Trans Embed Comput Syst (TECS) 6(4):22-es. doi:10.1145/1274858.1274860
4. Markus V, Michael K, Uwe Z (2004) Remoting patterns: foundations of enterprise, internet, and realtime distributed object middlware. Wiley, England, pp 130–133
5. Wang M, Cao J, Li J, Dasi SK (2008) Middleware for wireless sensor networks: a survey. J Comput Sci Technol 23:305–326
6. Cary P, Jorge C, John AM, Richard SP, Ivan V (2007) Introduction to web services. In: Cardoso J (ed) Semantic web services: theory, tools and applications. IGI Global, Hershey, pp 134–154, March 2007
7. Extensible Markup Language (XML) 1.0 (Fifth Edition). http://www.w3.org/TR/RECxml/. Accessed 29 Oct 2009
8. SOAP Version 1.2 Part 1: Messaging Framework (Second Edition). http://www.w3.org/TR/soap12-part1/. Accessed 29 Oct 2009
9. Web Services Description Language (WSDL) 1.1. http://www.w3.org/TR/wsdl. Accessed 29 Oct 2009
10. Borst WN (1997) Construction of engineering ontologies for knowledge sharing and reuse. Doctoral thesis of the University of Tweenty, Enschede, The Netherlands
11. Almeida MB, Bax MP (2003) An overview about ontologies: survey about definitions, types, applications, evaluation an building methods. Brasilia 32:3, 7–20. doi:10.1590/S0100-19652003000300002
12. Agilla: a mobile agent middleware for wireless sensor networks. http://www.cs.wustl.edu/mobilab/projects/agilla. Accessed 29 Oct 2009
13. Levis P, Culler DE (2002) Mat: a tiny virtual machine for sensor networks, architectural support for programming languages and operating systems. http://www.cs.berkeley.edu/pal/pubs/mate.pdf
14. TinyDB: a declarative database network. http://telegraph.cs.berkeley.edu/tinydb. Accessed 19 Oct 2009
15. TinyLime. http://lime.sourceforge.net/tinyLime/index.html. Accessed 19 Oct 2009
16. Iqbal M, Lim HB, Wang W, Yao Y (2009) A service-oriented model for semantics-based data management in wireless sensor networks. In: Proceedings of the 5th IEEE international workshop on heterogeneous wireless networks (HWISE 2009), May 2009
17. Rocha AR, Delicato FC, de Souza JN, Gomes DG, Pirmez L (2009) A semantic middleware for autonomic wireless sensor networks. In: Proceedings of the 2009 workshop on middleware for ubiquitous and pervasive systems (WMUPS '09), vol 389. ACM, Dublin, Ireland, New York, NY, pp 19–25, June 16–16. doi:10.1145/1551693.1551697
18. Java ME. http://java.sun.com/javame/index.jsp. Accessed Oct 2009

19. OWL Web Ontology Language. http://www.w3.org/TR/owl-features/. Accessed 29 Oct 2009
20. Gay D, Levis P, von Behren R, Welsh M, Brewer E, Culler D (2003) The nesC language: a holistic approach to networked embedded systems. In: Proceedings of the programming language design and implementation (PLDI), June 2003
21. TinyOs. http://www.tinyos.net/

CO_2 Gas Sensor Using Resonant Frequency Changes in Micro-Cantilever

S. Subhashini and A. Vimala Juliet

Abstract Our country India is facing some drastic changes in the climatic conditions due to the heating effect caused by various greenhouse gases. The most harmful gas among them is carbon dioxide and is increasing at an uncontrolled rate. This paper aims in finding out the quantity of the major polluting gas carbon dioxide. The gravimetric sensor works by absorbing the chemical in a CO_2 sponge, which alters the overall mass of the sensing element i.e., a cantilever, thereby its resonant frequency. Here a micro-cantilever beam is fabricated using selective alumina—silicate gel coatings on the surface to selectively absorb CO_2. As the gases are absorbed the mass increases and hence there is a change in resonant frequency. This change in frequency gives the measure of the quantity of gas present in that environment. The major expected advantage of this technique would be the repeatability of the sensor that is used.

1 Introduction

The term 'greenhouse gases' is used to refer to the gases present in the atmosphere which absorb the radiations and emit them within the thermal infrared range. These gases affect the temperature of the earth significantly, thus the "greenhouse effect" is the heating of the Earth due to the presence of greenhouse gases. The most prominent greenhouse gases in the Earth's atmosphere are water vapor, carbon dioxide, methane, nitrous oxide, ozone and CFCs.

The most harmful gases among them is carbon dioxide, as its content is increasing in the atmosphere day by day. Carbon dioxide is produced prominently by the

S. Subhashini (✉)
Sathyabama University, Chennai, India
e-mail: subhashinivivin@gmail.com

A. Vimala Juliet
Department of EIE, SRM University, Chennai, India

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_8,

respiration of animals, combustion of fossil fuels and geothermal processes, such as eruption of a volcano. Recent estimates reveal, that the concentration of carbon dioxide in the Earth's atmosphere has increased to 387 parts per million by volume. This gas features second in the greenhouse gases list, constituting 9 to 26 percent of greenhouse gases.

The average increase in the concentration of carbon dioxide in the Earth's atmosphere has been well documented. The graph below shows that the concentration increased from about 320 to over 380 ppm during the past 40 years. Locally, especially in urban areas, it is not uncommon for this concentration to be above 380 ppm due to emissions from fossil fuel combustion. The calibration value of 380 ppm for ambient outside air used with the CO_2 Gas Sensor is very close to these values. The ability of carbon dioxide to act as a greenhouse gas by absorbing increasing amounts of infrared radiation is a growing environmental concern. A graphical representation is shown in Fig. 1.

Miniaturization has gained enough significance as specified by Moores Law. Hence Micro-cantilevers are increasingly being used for various measurements. They are comparatively inexpensive, show fast response, have high sensitivity and are suitable for mass production using MEMS technology. Hence there is great interest in micro-cantilever based sensors.

This paper deals with an original methodology of measurement of the major polluting gas carbon dioxide. The gravimetric sensor works by absorbing the CO_2 in alumina—silicate gel, which alters the overall mass of the sensing element and thereby changing the resonant frequency of the cantilever used. This frequency shift is used to identify the quantity of CO_2 present in the atmosphere. Here a micro-cantilever beam is fabricated using selective polymer coatings on the surface to absorb CO_2.

A comparison among the various sensing techniques was done and the observations are tabulated (Table 1).

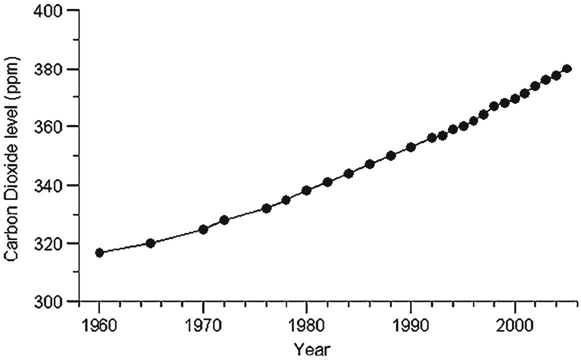


Fig. 1 Increasing global atmospheric carbon dioxide levels (source: earth system research laboratory, NOAA)

Table 1 Comparison among sensing techniques

Gas detection	Advantages	Disadvantages
Catalytic	Flammable gases	Contamination
Electrochemical	Toxic gases	Critical positioning
Point infrared	Physical technique and less sensitive to calibration errors	Critical positioning
Open path infrared	Flammable gases and toxic gases	High cost and path hindrance
Semiconductor	Mechanically robust	Susceptible to environmental changes
Thermal conductivity	Binary gas mixtures	High gas concentration
Paper tape	High sensitivity and selectivity for toxic gases	Extraction system

2 Sample Design and Fabrication

The sensing scheme is that a specific binding induced surface-stress causes bending of the sensing cantilevers. Present research integrates the SiO2 cantilever with alumina—silicate gel layer for the measurement of CO_2. The appearance formed would be a shown below in Fig. 2.

The important portion of this work is to selectively choose the coating that would be only sensitive to the CO_2 gas molecule of our interest. This is an interesting field of research for the chemist. They initially found out polymers that could be used on macro sized structures, that could adsorb the gas molecules. Later due to the inventions of micro sized structure they found that the physical parameters are altered and hence had to lookout for a more suitable adsorbing element. This lead to various findings like silicon carbide, but the repeatability of sensors became a question since the deadsorption of the gaseous molecules had to be carried out with some extra mechanical and electrical structures in fabrication process itself.

The surface area had to be increased for better sensitivity; this thought lead to the utilization of porous silicon carbide material. But this porous nature makes the design more complex as they are delicate and attempts to remove the molecules will lead to the rupture of porosity. Hence extra care should be taken in the designing aspect.

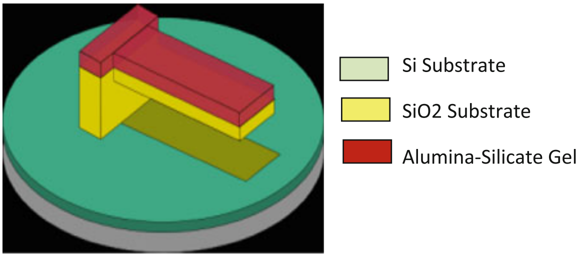


Fig. 2 3-D Image of an integrated cantilever

Due to the selectivity, the gas molecules are absorbed and hence there will be stress mounted due to the increase in weight and leads to a change in the resonant frequency of the cantilever too.

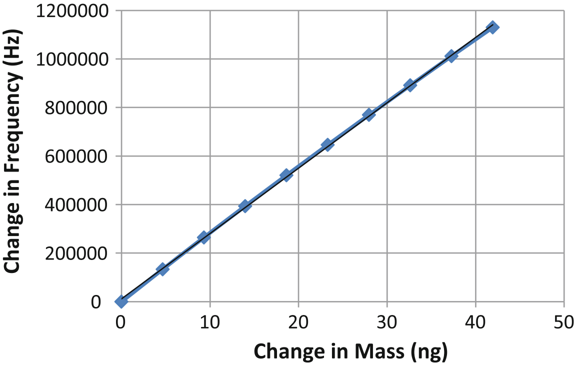
2.1 Cantilever Resonance Frequency Measurement for Chemical Gas Sensors

The principle used in this type of sensors is to literally catch gas molecules and to weigh them, measuring the shift in resonance frequency. The chemical sensor consists of two key components: a sensitive layer and the transducer (the cantilever). The sensing layer is the critical component and responsible for generating an initial signal from the interactions between the gas and the sensing layer. When the sensitive layer is exposed to gas, the cantilever mechanically responds by bending on the nanometer scale, because of the mass change and the gel swelling. The Molecular weight of CO_2 is 44 g/mol. The mass change is detected by measuring resonance frequency shifts while actuating the cantilever (dynamic mode). The additional mass load on the cantilever results in a decrease of the resonance frequency, which is detected. Hence if there is more deposition of the gases there will be more change in frequency and this is simulated using Intellisuite software and the results are consolidated and graphical represented in Fig. 3.

It is observed that as the change in mass increases there is a corresponding increase in change of frequency. We can hereby calculate the quantity of gas present in that environment from the resonant frequency we obtain. The frequency is related to mass by

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$$

Fig. 3 Change in mass versus frequency changes



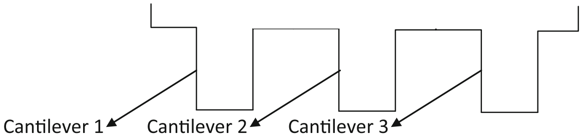


Fig. 4 Micro—cantilever array

Micro-Cantilever Array

Instead of having individual beams they can be combined as an array as shown in Fig. 4 so as to have a combined response.

An array of sensors can usually give a better amplified signal which would be of importance and also the selective coating could be varied to give us the quantitative details of N number of gas molecule detection. This could be achieved by just changing the selective coatings and hence obtaining the quantitative values for further signal conditioning circuits which could give an alarm.

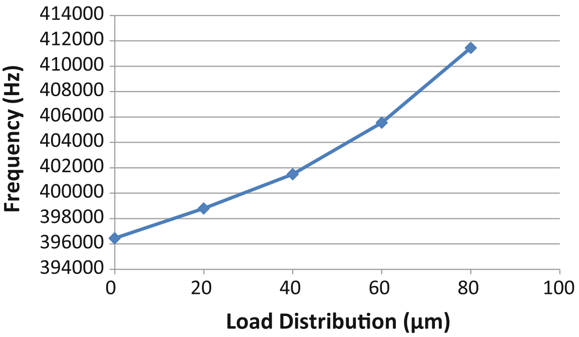
2.2 Sensitiveness Across the Beam

It would be of significance if we could get the maximum frequency change from the change in mass due to the gas being absorbed. For this reason a micro-cantilever of length: 100 μm, breadth: 20 μm and height 10 μm was considered.

The sensitive coatings of length 20 μm and thickness of 0.1 μm was integrated from the tip of the cantilever to the fixed end and the results were simulated using Intellisuite software and the results are consolidated and graphical represented in Fig. 5.

The sensitivity was found to be $4000/20 * 10^{-6}$ which is a very efficient one and hence this sensor helps us to identify gas molecules even when the quantity is very as ppb.

Fig. 5 Load distribution across the beam



3 Result and Conclusion

This paper thus brings out the original methodology of measurement of the major polluting gas like carbon dioxide. The absorption of the chemical by a sensitive layer alters the overall mass of the sensing element that varies the resonant frequency i.e an increase in mass leads to a decrease in resonant frequency, this change in frequency gives the mass of CO_2 gas molecules present in the atmosphere. We also observe that when the change in mass in more the resonant frequency change is also more. Finally a micro-cantilever beams sensitivity is observed. This paper could be developed using the same principle to identify various gas molecules.

References

1. Voiculescu I, Zaghloul M, McGill A (2003) Cantilever gas sensor. In: Proceedings of the IEEE international symposium on circuits and systems. Bangkok, Thailand, pp 25–28
2. Roukes M (2000) Solid-state sensor and actuator workshop, Hilton Head Island, SC, pp 367–370
3. Alexis M (2008) New CO_2 capturing material could make plants cleaner, ScienceDaily
4. Snow ES, Perkins FK, Houser EJ, Badescu SC, Reinecke TL (2005) Chemical detection with a single-walled carbon nanotube capacitor. Science 307(5717):1942–1945
5. Lavrik NV, Sepaniak MJ, Datskos PG (Jul. 2004) Cantilever transducers as a platform for chemical and biological sensors. Rev Sci Instrum 75(7):2229–2253
6. Ippolito SJ, Ponzoni A, Kalantar-Zadeh K, Wlodarski W, Comini E, Faglia G, Sberveglieri G (2006) Layered WO3/ZnO/36. LiTaO3 SAW gas sensor sensitive towards ethanol vapour and humidity. Sens Actuators B, Chem 117(2):442–450
7. Then D, Vidic A, Ziegler C (2006) A highly sensitive self-oscillating cantilever array for the quantitative and qualitative analysis of organic vapor mixtures. Sens Actuators B, Chem 117(1):1–9
8. Zinoviev K, Dominguez C, Plaza JA, Busto VJC, Lechuga LM (2006) A novel optical waveguide microcantilever sensor for the detection of nanomechanical forces. J Lightwave Technol 24(5):2132–2138

Part II
The Fourth International Conference on Networks & Communications (NETCOM-2012): Heterogeneous Wireless, WLAN and Mobile Networks

A Mechanism for Enhanced Performance of Chord DHT in Mobile Environment

Vu Thanh Vinh and Nguyen Chan Hung

Abstract With the growth of advanced networking technologies and capable terminals, it is very convenient for mobile operators to develop new service contents for mobile phones such as mobile Internet services, mobile television, and context-aware mobile services. Meanwhile, peer to peer (P2P) technology has made great success in multimedia services in the fixed networks such as Skype, PPlive, and Sopcast. Besides, demand for using P2P applications over PDAs and smartphone is increasing. So that, mobility in P2P networking has become a popular research area in recent years. However, mobile environment is different from the fixed networks in that it is limitation for process capability, memory, battery of mobile terminal, intermittent nature of wireless link and high churn rate of mobile users. In this paper, we focus on study drawbacks of Chord DHT algorithm in mobile environment. After that, we propose a mechanism for improving performance of Chord DHT algorithm based on the idea that is immediately stability of network topology when there is any node joining and leaving the network. Our simulation results show that our algorithm is better than previous studies about performance improvement of Chord DHT in mobile environment.

Keywords Distributed hash table · Chord DHT · DHT performance · High churn rate · Mobile peer to peer · Mobile peer to peer networks

V. T. Vinh (✉)
Thai Nguyen University of Information and Communication Technology,
Quyet Thang, Thai Nguyen city, Vietnam
e-mail: vtvinh@ictu.edu.vn

N. C. Hung (✉)
Vietnam Research Institute of Electronics, Informatics and Automation, Quan Thanh,
Hanoi, Vietnam
e-mail: hungnc@gmx.net

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_9,

1 Introduction

Recently, advances in mobile devices and communication technologies are growing with fast speed enabling fascinating opportunities for multimedia services to be widely deployed both on the Internet and in telecommunication networks. Mobile phones have used not only mainly for the communication purpose but also other purposes such as making of short videos, playing games and listening to music. Therefore, multimedia applications are spreading on mobile phone networks.

Whereas, P2P services such as file sharing services (e.g. BitTorrent, Gnutella) and communication services (e.g. Skype, PPlive, Sopcast) are very well known and popular for most of the fixed internet users. P2P technology offers an excellent media to distribute user created content. It is one of best choices for applications of multimedia communication by scalability, high reliability and balanced load.

Also, the mobile phone networks are moving towards the Internet architecture. Voice, video and signaling traffic is being transmitted over TCP/IP family protocols. 3G networks have used the Session Initiation Protocol (SIP) as their signaling protocol. Internet connectivity starts to be a standard feature in today's mobile phones, and users are accessing services (e.g. web, e-mail) from their cell phones regularly. So that mobile operators have started to show signs of interest to apply P2P technologies for mobile networks [1–3].

However, communication links is heavily affected by high packet loss rate and bandwidth fluctuations in mobile environment. In addition, with limitation for capabilities and resources, mobile devices have used various energy saving mechanisms try to save battery life by turning off unnecessary data transmission which shortens node online time [4, 5]. So that, mobile P2P services tend to join and leave very frequently due to the intermittent of communication link and user behavior, namely churn rate. In [6, 7] authors showed that churn rate is the most critical factor that affects the overall performance.

The rest of this paper is structured as follows: In Sect. 2, we give a brief reviews related studies. Section 3 describes our methodology in details. Analysis and evaluation simulation results of our mechanism are given in Sect. 4. Finally, Sect. 5 is the conclusions and future work.

2 Related Works

P2P network has evolved through three generations. The third generation P2P networks implements a mechanism called Distributed Hash Table (DHT) to better support scalability, load balancing and fault tolerance [8]. Since this model has proved to be effective in large scale networks, most modern P2P systems such as Chord DHT [9], Kelips DHT [10], and Tapestry DHT [11] are DHT-based. Chord DHT is a typical ring-based DHT, so the methodology and result obtain for Chord DHT can be apply to other ring-based DHT.

In [12], Ali Ghodsi introduces atomic ring maintenance to ensure correctness of lookups in the presence of joining and leaving nodes. The author showed some of the problems that can occur when nodes join and leave the DHT system. Ali Ghodsi only proposed several algorithms and proved it using analytical method without fully evaluates their performance.

S. Rhea et al. [13] using emulation of 1000 DHT nodes to evaluate several DHT mechanisms under churn. They focused on three issues: reactive versus periodic recovery from neighbor failure, the calculation of timeouts on lookup messages, and proximity neighbor selection. By introducing a timeout mechanism based on virtual coordinates, they tried to improve the mean latency. However, most of these works assume a normal scenario of file-sharing applications, without regard to the characteristics of mobile environments.

In our previous studied [14, 15], we proposed a mechanism which only modified the join process and consider leave as a failure since this is the common case in mobile wireless environment. In addition, we adapt a hybrid stabilization mechanism where proactive stabilization is activated when join activities happen and periodical stabilization still run in background. This mechanism is called Modify Ring Lock (MRL) mechanism. MRL mechanism is implemented into Chord DHT, namely MRLChord.

3 Methodology

3.1 Drawbacks of Chord DHT in Mobile Environment

In wired P2P networks, the join and leave of a node is mostly under user control and occur in rather low frequency, namely low churn rate (every 1–2 h or longer). Meanwhile, in mobile P2P applications, join and leave frequency of a mobile node is much higher, namely very high churn rate (every 120–600 s or lower), because of intermittent radio link, the battery saver mechanism of mobile terminal and user turn off or put his mobile terminal in standby mode [1, 4, 5].

Chord DHT uses a stabilization protocol running periodically in the background to update the successor list and the entries in the finger table. Successor list and finger table are stabilized separately. In case, when a node joins in network, Chord DHT does not immediately correct successor and predecessor pointers of all relevant nodes. Instead, Chord DHT only corrects some pointers and relies on periodical stabilization process to correct rest of pointers. This process is similar to node leave in network. It means that some pointers of relevant nodes are incorrect during the intervals between stabilization processes. Therefore, performance of Chord DHT decreases in mobile environment [14, 15].

3.2 ISR Mechanism

Our previous studies [14, 15] proposed MRL mechanism which only modified the join process and consider leave as a failure. It means that only join process of Chord DHT is immediately stabilized when join activities happen and periodical stabilization still run in background. Besides, our studies showed that Chord DHT has not any mechanism for solving leave process in fixed and mobile networks.

Therefore, we propose immediately stabilization ring (ISR) mechanism which modifies both the join process and leave process of Chord DHT based on previous studies [12, 14, 15]. The ISR mechanism's goal is to ensure that joins and leaves do not affect the functionality provided by the rest of the network. Particularly, ISR mechanism ensures that lookups to items in the hash table succeed while nodes are continuously joining and leaving the system.

The ISR mechanism includes two processes that are ISR for join process (ISRjoining) and ISR for leave process (ISRleaving). The ISRjoining process is depicted in Fig. 1, and ISRleaving process is depicted in Fig. 2.

In Fig. 1, when node r wants to join in network, it sends join request signal to a boot strap node. Join request is then forwarded to node q which is a successor of node r (1). Subsequently, node q sets predecessor pointer to r and then sends update response to node r for updating successor and predecessor (2). Upon receives update response, node r sets its successor pointer to q and predecessor pointer to p. Node r

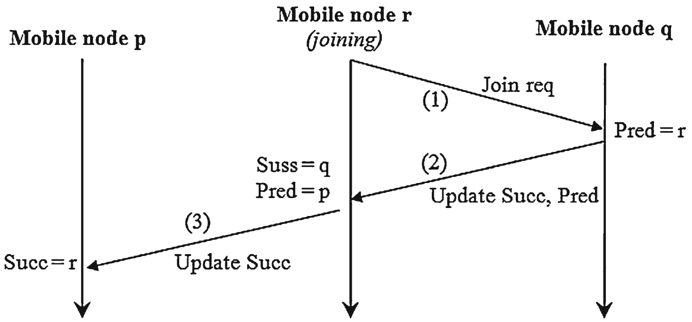


Fig. 1 Sequence diagram of the ISRjoining mechanism

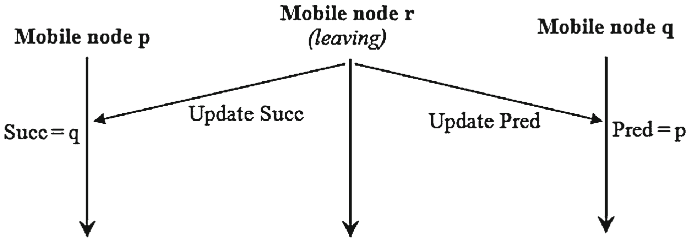


Fig. 2 Sequence diagram of the ISRleaving mechanism

then sends update pred message to node p to notify for updating new successor of node p (3). After node p receives this message, it sets its successor to node r. With this ISRjoining mechanism, stabilization not only runs periodically but also is triggered by update event, it means that the stabilization process starts immediately at the successful of the new joining node.

As can be seen from the Fig. 2, when node r wants to leave in network, update succ and update pred messages are sent by node r. Upon receives update response, node p sets its successor pointer to q, simultaneously node q sets its predecessor pointer to p.

In ISRleaving mechanism, stabilization not only runs periodically but also is triggered by update event, it means that the stabilization process starts immediately at the successful of node leave. With stabilization runs immediately when node join and node leave in network. Therefore, ISR mechanism solves the problem of inconsistent pointers by forcing intermediate stabilization in joining/leaving activities and reducing the inconsistency of pointers under failure. However, in case of node failure (without leave message), stabilization still runs periodically in background as basic Chord DHT [9].

4 Implementation and Evaluation

4.1 Implementation

We implement the ISR mechanism and stabilization algorithm in C++ code of OverSim [16], a new simulator for P2P network developed on top of OMNET++ framework [17], a popular graphical simulation. OverSim is an open-source overlay and peer-to-peer network simulation framework for the OMNeT++ simulation environment. OverSim contains several models for structured P2P system such as Chord, Kademlia, Pastry; and unstructured P2P system such as GIA. The interactive GUI and real-time messages of OverSim are extremely useful for debug and verification the new algorithms.

In this paper, we have implemented ISR mechanism into source code of Chord DHT in OverSim, namely ISRChord. For simulation leave event, we have added a parameter that is leaveNotify. We also implement two main functions which are ISRjoining and ISRleaving for solving node join process and node leave process of Chord DHT under mobile environment. Most study on performance evaluations of DHTs have focused on successful lookup ratio and lookup latency in under the same simulation conditions or unchanging network [6, 14, 15, 18]. So that, we also evaluate performance of ISRChord following these parameters. In our simulations, we are various network sizes from 100 nodes to 2000 nodes under various churn rates from 120 to 720 s. Table 1 list our simulation scenarios and the parameters of Chord DHT, MRLChord and ISRChord.

Table 1 Simulation scenarios and Chord parameters

Type	Parameters	Values
Simulation parameters	Node numbers	100, 500, 1000, 1500, 2000 (node)
	Churn rates	120, 180, 360, 540, 720 (s)
Common Chord parameters	Stabilize delay	10 s
	Fix fingers delay	20 s
	Successor list size	8
Chord modified parameters	LeaveNotify	0,5

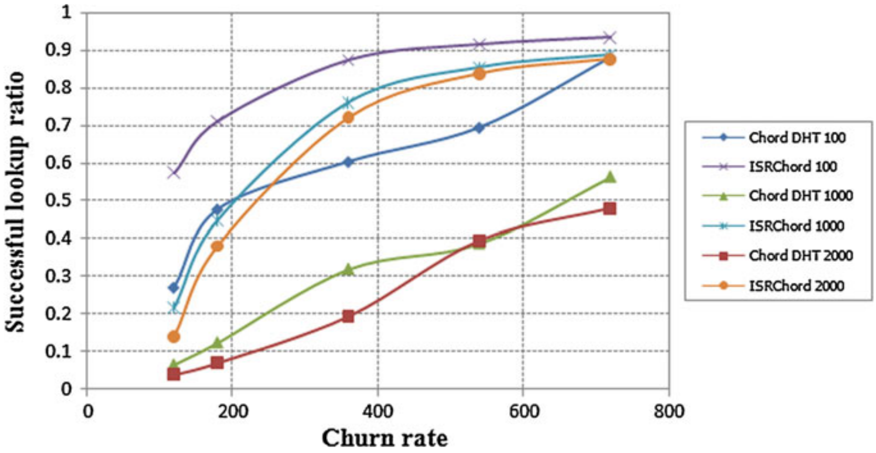


Fig. 3 Successful lookup ratio of Chord DHT and ISRChord operated in 100, 1000, 2000 node networks under various churn rates

4.2 Evaluation

In this section, we evaluate our new mechanism, ISRChord with Chord DHT [9] and MRLChord [14, 15]. As can be seen in Fig. 3, it is a comparison of successful lookup ratio of Chord DHT with ISRChord in network of 100, 1000 and 2000 nodes while churn rates varying from 120 to 720 s.

Figure 3 shows that ISRChord achieves successful lookup ratio much higher than Chord DHT in the same network size and under the same churn rate. For example, at churn rate of 180 s, successful lookup ratio of ISRChord of 100 nodes is 0.712, whereas successful lookup ratio of Chord DHT is only 0.476. This performance gain is due to the improvement in consistency of successor and predecessor pointers of nodes in ISRChord.

Besides, Fig. 3 has seen successful lookup ratio of both Chord DHT and ISRChord dramatically decrease when churn rate increase, and successful lookup ratio is closes to zero when churn rate is higher than 120 s.

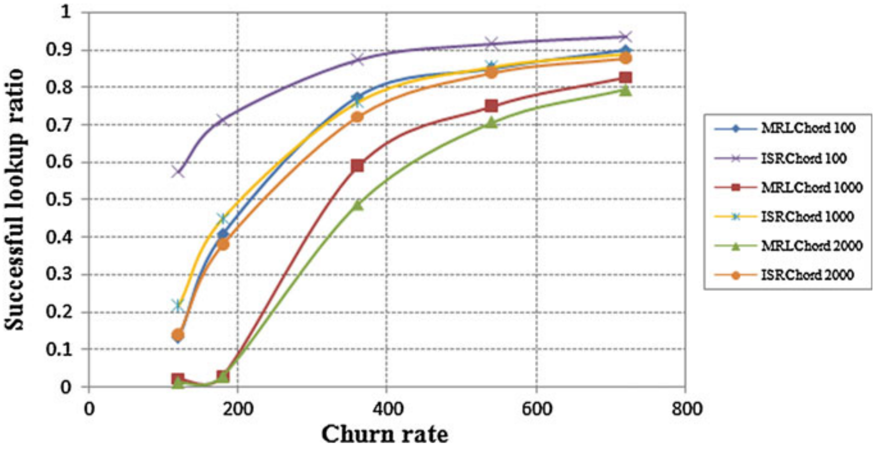


Fig. 4 Successful lookup ratio of MRLChord and ISRChord operated in 100, 1000, 2000 node networks under various churn rates

In Fig. 4, we can see that ISRChord achieves successful lookup ratio much higher than MRLChord in the same network size and under the same churn rate. Such as at churn rate of 360 s, successful lookup ratio of ISRChord of 1000 nodes is 0.761, whereas successful lookup ratio of MRLChord is 0.59. It means that ISR mechanism is better than MRL mechanism.

As can be seen from the Fig. 5, we show that successful lookup ratio of ISRChord under churn rate of 180, 360 and 720 s decreases when network size increase. This figure can be observed that ISRChord can achieve better performance in all case. For example, at churn rate of 360 s, successful lookup ratio of ISRChord of 1000 nodes is 0.761, whereas successful lookup ratio of Chord DHT is only 0.318. So that, performance of ISRChord is better than Chord DHT.

In Fig. 6, it shows that successful lookup ratio of ISRChord is higher than MRLChord in mobile environment. For example, at churn rate 360 s, successful lookup ratio of ISRChord of 1000 nodes is 0.761, whereas successful lookup ratio of MRLChord is 0.59. Specifically, successful lookup ratio of MRLChord is very low and closes to zero when churn rate is higher than 180 s. Whereas, successful lookup ratio of ISRChord is over 0.4 under same churn rate. So that performance of ISRChord is better than MRLChord.

In Figs. 7 and 8, we show that compare the parameter latency of successful lookup in Chord DHT and ISRChord networks of various sizes under various churn rates. As can be observed in these figures ISRChord has very little effect on latency under every churn rate, (slight increases in some cases) in all network sizing from 100 to 2000 nodes.

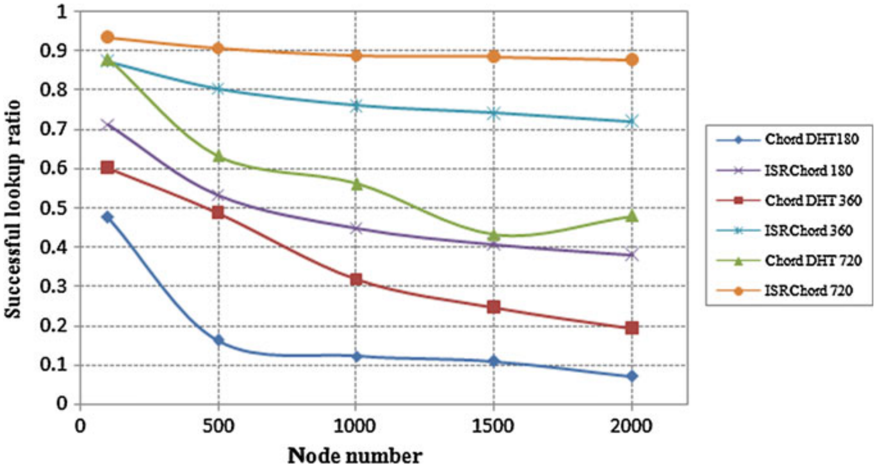


Fig. 5 Successful lookup ratio of Chord DHT and ISRChord operated in various network sizes under churn rate of 180, 360 and 720 s

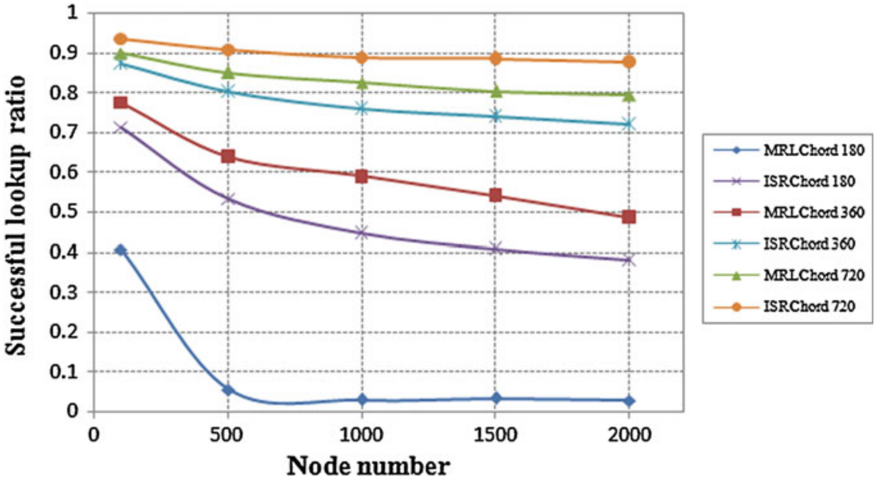


Fig. 6 Successful lookup ratio of MLRChord and ISRChord operated in various network sizes under churn rate of 180, 360 and 720 s

5 Conclusion and Future Work

In mobile environment, communication links is heavily affected by high packet loss rate and bandwidth fluctuations. mobile P2P services tend to join and leave very frequently, namely high churn rate. So that, we propose ISR mechanism and new stabilization strategy to ensure the correctness of node pointers under high churn rate.

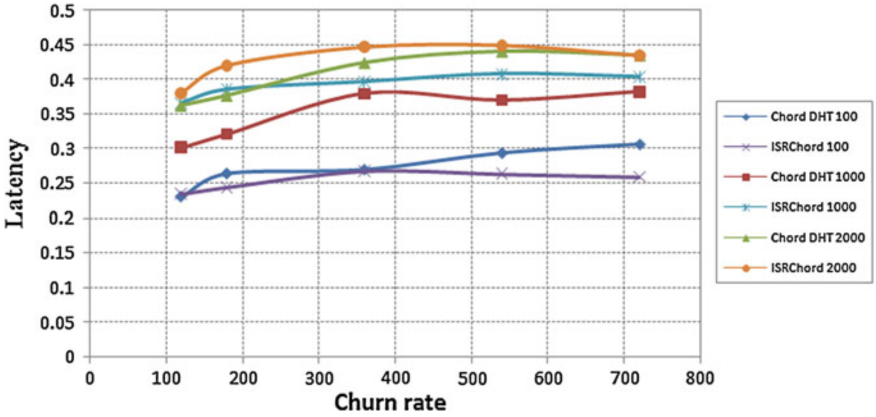


Fig. 7 Latency of Chord DHT and ISRChord operated in 100, 1000, 2000 node networks under various churn rates

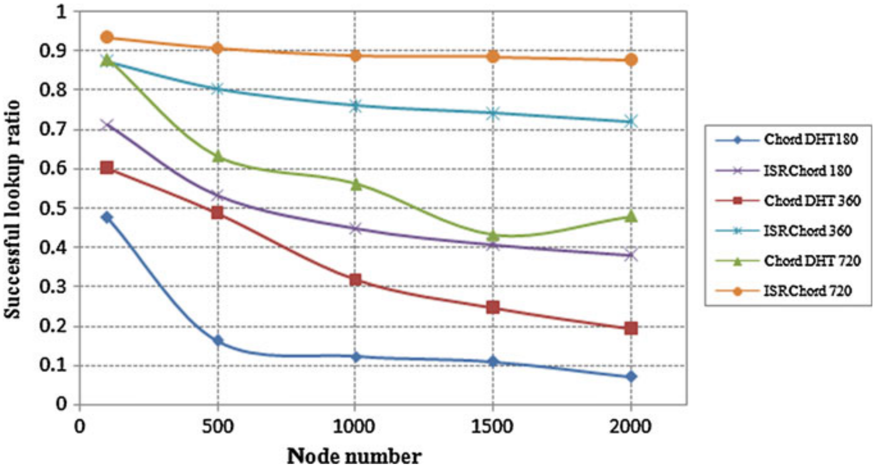


Fig. 8 Latency of Chord DHT and ISRChord operated in various network sizes under churn rate of 180, 360 and 720 s

We also successfully implement the ISR mechanism into Chord DHT in OverSim simulator. The simulation results show significant improvement in successful lookup ratio of ISRChord over Chord DHT under various conditions of churn rate and network size. However, ISRChord does not affect much on other aspects of Chord DHT such as lookup latency. Besides, we also compare ISRChord with MLRChord that is previous our mechanism [14, 15]. It shows that performance of ISRChord is better than performance of MRLChord in mobile environment.

In the next phase of our study, we will study the feasibility of this solution and develop ISR mechanisms for mobile P2P applications in networks.

Acknowledgments The authors express the VLIR project key program ZEIN2011Z099 for supporting this research.

References

1. Yrjö R (2005) Mobile peer-to-peer in cellular networks. In: Proceedings of HUTT-110.51 seminar on internetworking, Helsinki Institute of Technology, pp 4–26
2. Beijar N et al (2005) Mobile peer-to-peer content sharing services in IMS. In: International conference on telecommunication systems, modelling and analysis, USA
3. Wang I, Saxlund K (2007) Peer2Me—rapid application framework for mobile peer-to-peer applications. In: International symposium on collaborative technologies and systems
4. Hung NC, Vinh VT et al (2009) Challenges in the development of mobile P2P applications and services. J Inf Technol Commun, VietNam ministry of information and communications
5. Motta R, Pasquale J (2010) Wireless P2P: problem or opportunity?. In: The second international conference on advances in P2P systems
6. Li J et al (2004) Comparing the performance of distributed hash tables under churn. In: LNCS 3279, Springer-Verlag, Berlin, Heidelberg, pp 87–99
7. Li B et al (2007) An empirical study of the coolstreaming + system. J Sel Area Commun 25(9):1627
8. Shen X et al (2010) Handbook of peer-to-peer networking. Springer, New York. ISBN 978-0-387-09750-3
9. Stoica I et al (2001) Chord: a scalable peer-to-peer lookup service for internet applications. In: Proceedings of the 2001 ACM SIGCOMM conference, pp 149–160
10. Gupta I et al (2003) Kelips: building an efficient and stable P2P DHT through increased memory and background overhead. In: Proceedings of the 2nd IPTPS
11. Zhao BY et al (2004) Tapestry: a resilient globalscale overlay for service deployment. IEEE J Sel Areas Commun 22(1):41–53
12. Ghodsi A (2006) Distributed k-ary system: algorithms for distributed hash tables, PhD dissertation, KTH-Royal Institute of Technology
13. Rhea S et al (2004) Handling churn in a DHT. In: Proceedings of the 2004 USENIX technical conference
14. Chan HN, Vinh VT et al (2009) Performance improvement of Chord distributed hash table under high churn rate. In: Proceeding of international conferences on advanced technologies for communication, Vietnam
15. Hung NC, Vinh VT et al (2010) Improvement of successful lookup ratio of chord distributed hash table (DHT) in wireless communication environment. In: Proceedings of international conference on advanced technologies for communications (ATC), Vietnam
16. OverSim. http://www.oversim.org
17. Omnet ++. http://www.omnetpp.org/
18. Ou Z et al (2010) Performance evaluation of a Kademlia-based communication-oriented P2P system under churn. Comput Netw 54:689–705

A Hybrid Model of CLMS and ACLMS Algorithms for Smart Antennas

Y. Rama Krishna, P. E. S. N. Krishna Prasad, P. V. Subbaiah and B. Prabhakara Rao

Abstract Smart Antenna is a device that enables to steer and modify an arrays beam pattern to enhance the reception of a desired signal, while simultaneously suppressing interfering signals through complex weight selection. The weight selection process is a complex method to get low Half Power Beam Width (HPBW) and Side Lobe Level (SLL). The aim of this task is to minimize the noise and interference effects from external sources. This paper presents a Hybrid based model for Smart Antennas by combining CLMS and Augmented CLMS algorithms. Since CLMS and ACLMS models have their own pros and cons in the process of adaptive beam forming, Hybrid model results a better convergence towards desired signal, Low HPBW and low SLL in the noisy environment.

Keywords Hybrid model of CLMS and ACLMS · Adaptive array · Adaptive beamforming · Smart antennas · Wireless sensor networks · Complex least mean square (CLMS) · Augmented CLMS (ACLMS) · Side lobe level (SLL) · Half power beam width · Error convergence rate

1 Introduction

Wireless cellular networks [1, 2] are fast growing technology in the current world and this trend is likely to continue for several years. The advancements in radio technology enable novel and improved services in the cellular systems. Current wireless services include transmission of voice, fax, and Multimedia applications and so on. Multimedia services like video-on demand and internet access needs

Y. Rama Krishna (✉) · P. E. S. N. Krishna Prasad
PVP Siddhartha Institute of Technology, Vijayawada, India
e-mail: ramakrishna.yarlagadda@gmail.com

P. V. Subbaiah
Amrita Sai Institute of Science and Technology, Vijayawada, India

B. Prabhakara Rao
JNT University Kakinada, Kakinada, India

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_10,

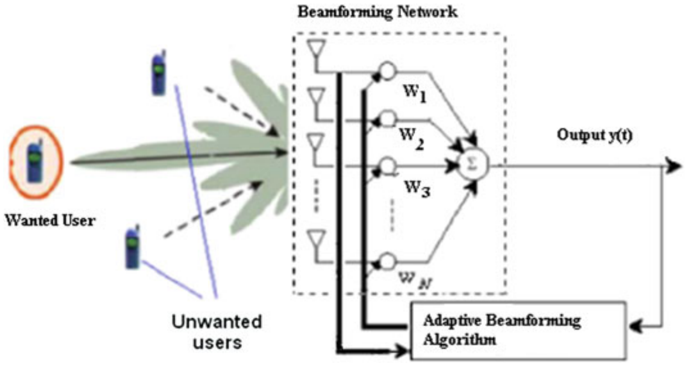


Fig. 1 Basic smart antenna system

of environments, spanning dense urban, suburban, and rural areas. Mobility needs varying must also be addressed.

Smart Antennas (SA) [3, 4] consists of an array of antenna elements with signal processing capability that optimizes the radiation and reception of a desired signal dynamically. SAs can place nulls in the direction of interferers via adaptive updating of weights linked to each antenna element. SAs [5], thus cancel out most of the co-channel interference resulting the better quality of reception and lowered dropped calls. SAs can also track the user within a cell via direction of arrival algorithms (Fig. 1).

The Smart antennas [3, 6] perform spatial filtering, which makes it possible to receive energy from a particular direction, while simultaneously blocking it from another direction. This property makes smart antennas as an effective tool in detecting and locating radiation from other sources. That means, the design and development of the efficient models to this task for real time optimization is a current problem. The control unit of the Smart Antenna is normally realized with a Digital Signal Processing (DSP) unit. Based on certain inputs, the DSP controls radiation parameters of the antenna to optimize the communication link. The following figure shows the basic model of Smart Antenna (SA).

2 Complex Adaptive Algorithms

A Complex valued Neural Network [7–9] is an artificial neural network, consists of complex valued input signals, weights, threshold values and/or signal functions. Such kind of models must be needed for solving the problems in the field of signal processing. In the signal processing, signals are complex valued and processing of such signals requires the implementation of new complex valued neural processing models. One of the most important characteristics of the complex valued neural models is processing of linear complex signals of the smart antennas. In smart antennas, signals from different sources or interferers are to be processed before orienting the main beam direction of the antenna array. In this context identifying the angle of arrival of the desired signal is very important. More over the Half Power Beam Width (HPBW) and Side Lobe Level (SLL) of the array radiation pattern must be as small as possible to avoid the interference.

In this work Complex Least Mean Square (CLMS) and Augmented Complex Least Mean Square (ACLMS) algorithms [10, 11] are considered as complex valued neural networks and these models are applied on complex signals of Smart Antenna System.

Least Mean Square (LMS) algorithm is a fundamental gradient based algorithm introduced by Widrow and Hoff in 1959 that estimates the gradient vector from the available data. This algorithm is an iterative method that leads to Minimum Mean Square Error (MSE), but this is a simple model which cannot process complex data with more noise. In the analysis of LMS, its convergence is slow if the Eigen values are widely spread and the convergence of LMS directly depends on the Eigen structure. The convergence time of LMS can be exceedingly long and highly data dependent when the Eigen values of the covariance matrix differs.

In order to process such complex signals, we choose a variant of LMS models such as Complex Least Mean Square (CLMS) and Augmented Complex Least Mean Square (ACLMS) algorithms and also a combination of these two models as Hybrid. This paper presents a performance analysis of these models on complex signals and the results are discussed in Sect. 3.

2.1 Complex Least Mean Square Algorithm (CLMS)

Complex Least Mean Square (CLMS) [11–13] algorithm was introduced by Widrow et al. in 1975, which benefits from the robustness and stability of the LMS and enables the simultaneous processing of complex signals. This algorithm performs stochastic gradient decent in complex domain statistics that enables better modelling of complex data and produce effective outcome. The basic algorithm of CLMS is as follows:

1. The instantaneous estimates

$$E[|e^2(k)|] \rightarrow 1/2|e^2(k)| \tag{1}$$

$$E[x(k)x^H(k)] \rightarrow x(k)x^H(k) \tag{2}$$

$$E[x(k)d(k)] \rightarrow x(k)d(k) \tag{3}$$

2. The ‘stochastic’ cost function

$$J(k) = 1/2|e(k)|^2 \tag{4}$$

3. Weight vector update

$$w(k+1) = w(k) - \mu\nabla_w J(w) \tag{5}$$

4. The gradient of the cost function
The gradient of the cost function with respect to the complex valued weight vector $w(k) = w_r(k) + jw_i(k)$ can be expressed as

$$\nabla_w J(k) = \nabla_{W_r} J(k) + j\nabla_{W_i} J(k) = \frac{\partial J(k)}{\partial w_r(k)} + j\frac{\partial J(k)}{\partial w_i(k)} \tag{6}$$

5. The output error is given by

$$e(k) = d(k) - x^T(k)w(k) \tag{7}$$

6. The stochastic gradient adaptation for the weight vector can be expressed as

$$w(k+1) = w(k) + \mu_1 e(k)x^*(k), w(0) = 0 \tag{8}$$

7. This output of the complex least mean square (CLMS) algorithm is computed as

$$y = x^H(k)w(k) \longrightarrow w(k+1) = w(k) + \mu_1 e(k)x(k) \tag{9}$$

2.2 *Augmented Complex Least Mean Square Algorithm (ACLMS)*

The ACLMS [10, 11, 13] algorithm has the same generic form as the standard CLMS, it is simple to implement, yet it takes into account the full available second-order statistics of complex valued inputs (non circularity) in the domain of adaptive beam forming that utilizes the second order statistical information. This is achieved based on some advancement in complex statistics with the use of widely linear modelling. So this model is also called as widely Linear LMS. The advantages of ACLMS over CLMS are as follows:

- In blind source separation, it is able to deal with more sources than observations,
- Improved signal recovery in communication modulation schemes,
- Improved direction of arrival estimation in augmented array signal processing,
- The analysis of augmented signal processing algorithms benefits from special matrix structures which do not exist in standard complex valued signal processing.

The basic algorithm of ACLMS is as follows:

1. The output of ACLMS is

$$\begin{aligned} y(k) &= \sum\nolimits_{n=1}^{N} [h_n(k)z(k-n) + g_n(k)z^*(k-n)] \Leftrightarrow y(k) \\ &= h^T(k)z(k) + g^T(k)z^*(k) \end{aligned} \tag{10}$$

2. Weight updation:

$$\Delta w_n(k) = \mu_2 \nabla w_n J(k) = \mu_2 \frac{\partial Jx}{\partial w_n(k)} = \mu_2 \left(\frac{\partial J(k)}{\partial w_n^r(k)} + j \frac{\partial J(k)}{\partial w_n^i(k)} \right) \tag{11}$$

where $w_n(k) = w_n^r(k) + jw_n^i(k)$ is a complex weight and μ_2 is the learning rate, a small positive constant.

$$h(k+1) = h(k) + \mu_2 e(k)z^*(k) \tag{12}$$

$$g(k+1) = g(k) + \mu_2 e(k)z(k) \tag{13}$$

3. The augmented weight vector

$$w^a(k) = [h^T(k), g^T(k)]^T \tag{14}$$

4. The final form of ACLMS as follows:

$$w^a(k+1) = w^a(k) + \mu_2 e(k)z^{a*}(k) \tag{15}$$

where the 'augmented' instantaneous error is

$$e(k) = d(k) - z^a T(k) w^a(k) \tag{16}$$

2.3 Hybrid Model

In the Hybrid Model, CLMS and ACLMS have been combined since each algorithm has its own merits and demerits in the process of adaptive beam forming of signals in Smart Antennas. In CLMS even though the HPBW is low, the SLL is high which is not desired and moreover the array output signal convergence towards the desired signal is poor that can be discussed in Sect. 3 (Fig. 2 and Table 1). Similarly in ACLMS even though the SLL is low, the HPBW is high which is not desired and moreover the array output signal convergence towards the desired signal is also poor that can be discussed in Sect. 3 (Figs. 2, 3 and Table 1).

In the analysis of CLMS and ACLMS with experimental results we proposed a model of combining these two algorithms as hybrid which is often referred as Hybrid of CLMS and ACLMS. The hybrid algorithm is as follows:

$$y_{CLMS}(k) = x^H(k)w_{CLMS}(k) \rightarrow w_{CLMS}(k+1) = w(k) + \mu_1 e(k)x(k) \tag{17}$$

$$y_{ACLMS}(k) = h^T(k)z(k) + g^T(k)z*(k) \tag{18}$$

$$h(k+1) = h(k) + \mu_2 e(k)z^*(k) \tag{19}$$

$$g(k+1) = g(k) + \mu_2 e(k)z^*(k) \tag{20}$$

$$e_{CLMS}(k) = d(k) - x^T(k)w_{CLMS}(k) \tag{21}$$

$$e_{ACLMS}(k) = d(k) - z^{aT}(k)w_{ACLMS}(k) \tag{22}$$

$$\mathrm{Y_{Hybrid}} = \lambda * \mathrm{Y_{CLMS}}\,(\mathrm{k}) + (1-\lambda) * \mathrm{Y_{ACLMS}}\,(\mathrm{k}) \tag{23}$$

$$\mathrm{e_{Hybrid}}\,(\mathrm{k}) = \mathrm{x} - \mathrm{Y_{Hybrid}}\,(\mathrm{k}) \tag{24}$$

$$\mathrm{W_{Hybrid}}\,(\mathrm{k}+1) = \lambda\,(\mathrm{k})\,\mathrm{W_{CLMS}}\,(\mathrm{k}) + (1-\lambda)\,\mathrm{W_{ACLMS}}\,(\mathrm{k})^* \tag{25}$$

$$\lambda\,(\mathrm{k}+1) = \lambda\,(\mathrm{k}) + \mu 3 * \mathrm{real}(\mathrm{e_{Hybrid}}\,(\mathrm{k}) * (\mathrm{Y_{CLMS}}\,(\mathrm{k}) - \mathrm{Y_{ACLMS}}\,(\mathrm{k})) \tag{26}$$

In Algorithm 1, 2, 3 [7, 11, 12, 14], some notations have been considered for step size parameters such as $\mu 1$ for CLMS, $\mu 2$ for ACLMS, $\mu 3$ for Hybrid and λ for mixing parameter.

Also we have framed certain constraints for constructing the Hybrid model by the combination of CLMS and ACLMS as

1. When we are considering the μ values for these algorithms, we need to follow the conditions for step size parameters. $\mu_3 > \mu_1 > \mu_2$ and $\mu 3 = 0.8$
2. As $\mu 1$ and $\mu 2$ are considered the same values, CLMS and ACLMS gives the best performance for HPBW and SLL but in the case of Hybrid it cannot.
3. As μ values are reduced to low levels the convergence of array output towards desired signal for CLMS and ACLMS is satisfactory but when the μ values are low, the convergence towards desired signal will be too poor, which causes over damped condition.
4. It is also observed that the Hybrid algorithm giving optimum performance in terms of lower values of N. i.e. number of array elements. Hence the Hybrid model reduces the array length of Smart Antenna which reduces the system complexity.

3 Result Analysis

In the process of adaptive beam forming signals [4, 5], an input signal $x_s(k) = \cos(2wt)$ with frequency 1 kHz has been considered along with noise. Using CLMS and ACLMS, the Half Power Beam width and SLL are computed from the array factors plotted with the normalized weights and presented in the Table 1. The same data is plotted in Fig. 3. From these two algorithms, HPBW and SLL are computed individually that gives good results at N = 9 and $\mu = 0.002$ for CLMS and N = 8, $\mu = 0.002$ for ACLMS. The analysis of these two algorithms is presented in Fig. 3.

Figure 2 shows signal convergence of CLMS and ACLMS and Fig. 3 shows Array factor comparison of CLMS and ACLMS for SLL and HPBW at different N and μ values.

This paper proposed a novel model by combining CLMS and ACLMS algorithms as hybrid adaptive beamforming of signals. The hybrid model performs a best outcome of array elements when compared with CLMS and ACLMS algorithms. Also the Hybrid model produces low HPBW and SLL than the existing models and gives better convergence of the desired signal in the noise environment which is presented Table 2 and also presented in Figs. 4, 5, 6 and 7. In Figs. 4, 5 and 6, N is the number of Array elements, $\mu 1$ is the Step Size parameter of CLMS Algorithm, $\mu 2$ is the Step Size parameter of ACLMS Algorithm and $\mu 3$ is the Step Size parameter of Hybrid Algorithm.

Figure 4 presents the analysis of array outputs of CLMS, ACLMS and Hybrid models towards convergence with desired signal at N = 8, $\mu 1 = 0.2$, $\mu 2 = 0.005$

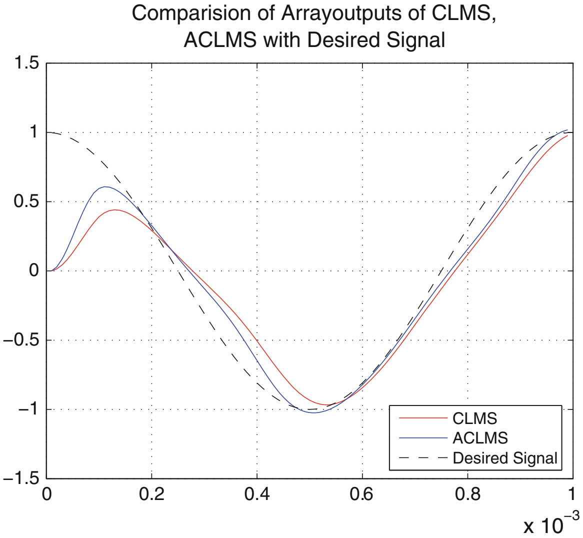


Fig. 2 Signal convergence comparison for CLMS and ACLMS algorithms

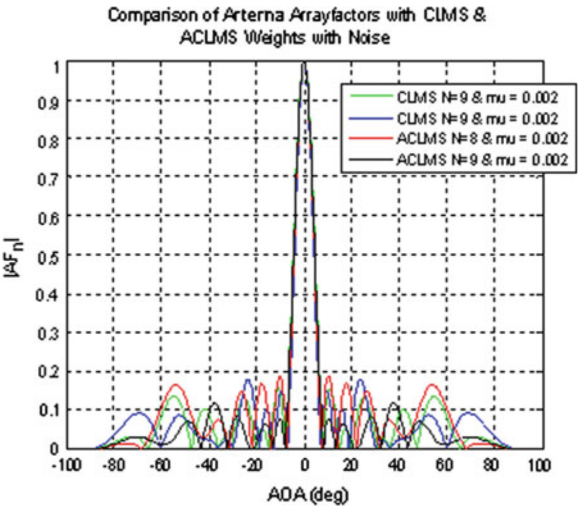


Fig. 3 Comparison of best performance by CLMS and ACLMS weights in noisy environment

Table 1 Performance analysis of CLMS ACLMS algorithms with noise

Sl. No	Weights used	No. of array elements(N)	Step size parameter (μ)	HPBW (%)	No. of symmetrical side lobes	SLL
1	CLMS	8	0.002	6.6	7	0.155
2	CLMS	9	0.002	6.1	6	0.1781
3	ACLMS	8	0.002	6.4	6	0.1873
4	ACLMS	9	0.002	6.5	6	0.1189

& $\mu 3 = 0.8$. Figure 5 presents the analysis of array outputs of CLMS, ACLMS and Hybrid models towards convergence with desired signal at N = 9, $\mu 1 = 0.1$, $\mu 2 = 0.002$ & $\mu 3 = 0.8$. Figure 6 presents the analysis of array outputs of CLMS, ACLMS and Hybrid models towards convergence with desired signal at N = 8, $\mu 1 = 0.1$, $\mu 2 = 0.002$ & $\mu 3 = 0.8$.

Table 2 Results analysis of CLMS, ACLMS and Hybrid model at best values of N and μ for producing low HPBW and SLL

Sl. No	N	$\mu 1$ (CLMS)	$\mu 2$ (ACLMS)	$\mu 3$ (HYBRID)	HPBW			SLL		
					CLMS	ACLMS	HYBRID	CLMS	ACLMS	HYBRID
1	8	0.2	0.005	0.8	5.8	7.3	7.2	0.525	0.555	0.0810
2	9	0.1	0.002	0.8	16.6	6.3	9.9	1.095	0.1312	0.0204
3	8	0.1	0.002	0.8	–	6.9	10.1	–	0.2252	0.0192

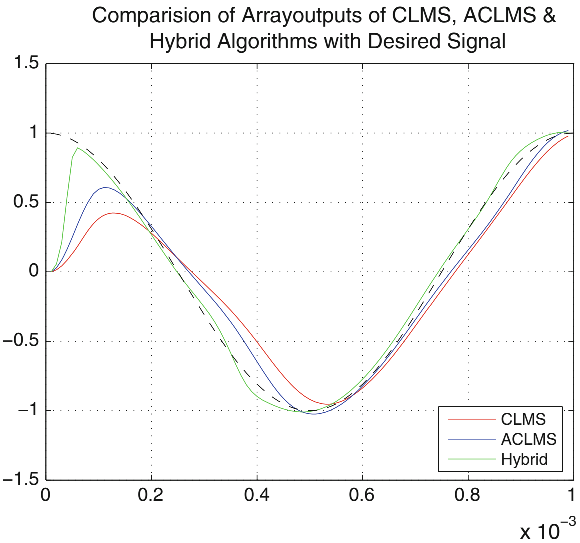


Fig. 4 Performance of array outputs of CLMS, ACLMS and Hybrid models towards convergence with desired signal at N $= 8$, $\mu 1 = 0.2$, $\mu 2 = 0.005$ & $\mu 3 = 0.8$

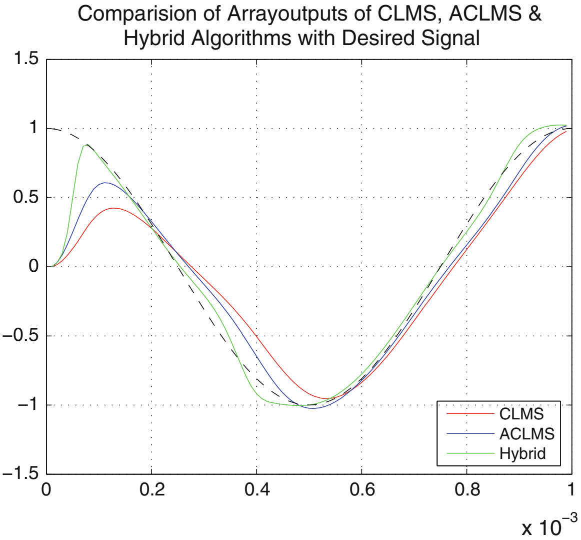


Fig. 5 Performance of array outputs of CLMS, ACLMS and Hybrid models towards convergence with desired signal at N $= 9$, $\mu 1 = 0.1$, $\mu 2 = 0.002$ & $\mu 3 = 0.8$

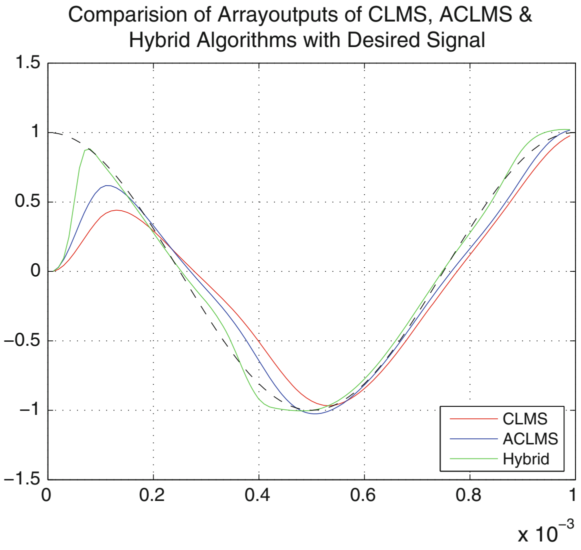


Fig. 6 Performance of array outputs of CLMS, ACLMS and Hybrid models towards convergence with desired signal at N = 8, $\mu 1 = 0.1$, $\mu 2 = 0.002$ & $\mu 3 = 0.8$

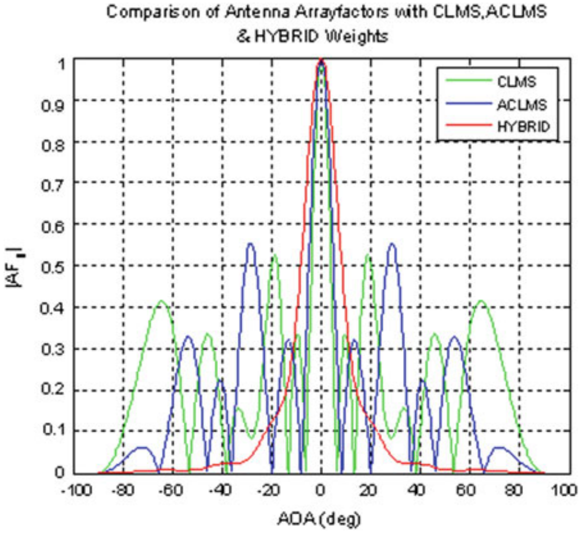


Fig. 7 Performance of CLMS, ACLMS and Hybrid models in noisy environment

In Fig. 7 the best performance of array factors of CLMS, ACLMS and Hybrid models is presented at N = 8, $\mu 1 = 0.2$, $\mu 2 = 0.005$ & $\mu 3 = 0.8$. CLMS gives low HPBW = 5.8, ACLMS 7.3 and Hybrid 7.2 but SLL for CLMS 0.525, ACLMS 0.555 and Hybrid 0.0810. From this data CLMS results better HPBW but SLL is too high this cannot be preferred in mobile networks. As per analysis of the array factors, the hybrid model returns good low SLL value and optimal convergence towards the desired signal. So the proposed model gives better performance than CLMS and ACLMS for the parameters N, μ and optimal convergence.

4 Conclusion

From the observations and the analysis of CLMS, ACLMS and Hybrid Models for adaptive beamforming in Smart Antennas, Hybrid model converges approximately near to the desired signal and retains optimum values for HPBW and SLL. So the Hybrid model is the optimal model of CLMS and ACLMS rather than the individuals in Smart Antenna System. From the observations of Sect. 3, $\mu 3$ is bounded at 0.8 and $\mu_1 > \mu_2$. The future scope of this task is to design and develop an efficient stochastic based model that improves the overall performance of the Smart Antenna System.

References

1. Winters JH (1998) Smart antennas for wireless systems. IEEE Pers Commun 5(1):23–27
2. Godara LC (1997) Applications of antenna arrays to mobile communications. I. Performance improvement. IEEE Proc Feasibility Consid 85(7):1031–1060
3. Stevanovic I, Skrivervik A, Mosig JR (2003) Smart antenna systems for mobile communications. Laboratoire d'Electromagnetisme et d'Acoustique Ecole Polytechnique Federale de Lausanne (2003)
4. Smart Antenna Systems Tutorial, The international engineering consortium. www.iec.org
5. Smart Antennas Beamforming Tutorial. www.altera.com
6. Haykin S, Kailath T (2009) Adaptive filter theory, 4th edn. Pearson Education
7. Mandic D, Vayanos P, Boukis C, Jelfs B, Goh SL, Gautama T, Rutkowski T (2007) Collaborative adaptive learning using hybrid filters. In: ICASSP 2007, vol 3, pp 921924 (2007)
8. Haykin S, Li L (1995) Nonlinear adaptive prediction of non-stationary signals. IEEE Trans Signal Process 43(2):526–535
9. Hirose A (2003) Complex valued neural networks: theories and applications. World Scientific Publications, Singapore
10. Javidi S, Pedzisz M, Goh SL, Mandic DP (2008) The augmented complex least mean square algorithm with application to adaptive prediction problems. In: Proceedings of the 1st IARP workshop on cognitive information processing, pp 54–57 (2008)
11. Mandic DP, Xia Y, Douglas SC (2010) Steady state analysis of the CLMS and augmented CLMS algorithms for non-circular complex signals. In: Proceedings of ASILOMAR, pp 1635–1639 (2010)
12. Mandic DP, Xia Y, Syad AH (2011) An adaptive diffusion augmented CLMS algorithm for distributed filtering of non-circular complex signals. IEEE Signal Process Lett 18(11):659–662

13. Mandic DP, Goh VSL (2009) Complex valued nonlinear adaptive filters noncircularity, widely linear and neural models. Wiley, Chichester
14. Mandic DP, Javidi S, Souretis G, Goh SL (2007) Why a complex valued solution for a real domain problem. In: Proceedings of the 17th IEEE signal processing society workshop on machine learning for, signal processing (2007)
15. Schreier PJ, Scharf LL (2003) Second-order analysis of improper complex random vectors and processes. IEEE Trans Signal Process 51(3):714–725

Novel Protection from Internal Attacks in Wireless Sensor Networks

Xu Huang, Muhammad Ahmed and Dharmendra Sharma

Abstract Due to wireless sensor networks are easy and rapid deployed, low cost, low power, self-organized, cooperatively collect the environmental information and realize the integration of the physical world and communication network, they become part of our daily life. However, security threats to WSNs become increasingly diversified and preventions are getting harder and harder due to the open nature of the wireless medium. For example, an adversary can easily eavesdrop and replay or inject fabricated messages. Different cryptographic methods are very limited. This is because of internal attack, such as node compromise, becomes another major problem that is different from traditional WSN security problem as it allows an adversary to enter inside the security perimeter of the network. This situation raised a serious challenge for the security of WSNs. In this paper we are investigating internal attacks of wireless sensor networks, with an example of multi-hop and a single sinker, by which we present our novel algorithm with controllable robust protecting from internal attacks of a wireless sensor network. The final experimental works showed that the proposed algorithm does work well at the designed level.

Keywords Internal attack · Robust protection · Resiliency of WSN · Wireless sensor networks · Sensor optimum deployment · Network security

1 Introduction

Wireless sensor networks (WSNs) and their applications are becoming part of our daily life, they have a great advantage for various applications in our real life [1], such as habitat monitoring, battlefield surveillance, intelligent agriculture, home

X. Huang · M. Ahmed (✉) · D. Sharma
Faculty of Information Sciences and Engineering, University of Canberra,
Bruce, Australia
e-mail: muhammad.ahmed@canberra.edu.au

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_11,

automation, etc. However, the properties of WSN inevitably have the natures that are extremely restricted by their resources, including energy, memory, computing complexity, bandwidth, and communication capacity. Normally the base station is a more powerful node, which can be linked to a central station via satellite or internet communication to form a network. There are many deployments for wireless sensor networks depending on various applications, such as, environmental monitoring, volcano detection [1–3], distributed control systems [4], agricultural and farm management [5], detection of radioactive sources [6], and computing platform for tomorrows' internet [7]. However, the open nature of the wireless medium therefore offers chances for an adversary to easily eavesdrop information from the sensors, or actively do something such as replay or inject fabricated messages. It is well known that for the protection from the some WSNs attacks, various cryptographic methods are widely used but sometimes are not very efficient and effective [8–10]. Moreover, because of WSN deployments are in open states and possibly used in hostile environments, attackers can easily lunch "denial-of-service" (DoS) attacks, cause physical damage to sensors or capture them to extract sensitive data from those sensors, such as identities, encryption keys, address, and other privacy data, etc. Recently internal attacks attracted great attentions to the people who have been working in the fields as it allows an adversary to enter inside the security perimeter of the network. For instance, form a node of a WSN, so-called compromised, the attack can produce internal attack, such as Sybil attacks, node replication or black-grey-worm-sink holes. As mentioned above that sometimes cryptography to secure routing functionalities is inappropriate against the aforementioned internal attacks, this is due to those attacks can introduce false topological, neighborhood, control, and routing information. It also may just simply drop message as a black hole. So far, there is limited research literature in investigating and analyzing against this type attacks. In this paper, we are focusing on investigating those internal attacks of wireless sensor networks, by which we can show with our novel algorithm that is under some fixed resiliency degree the targeted WSN can be controlled with the fixed resiliency by the design. Therefore, the designed network will always work within the required safe level.

This paper consists of five sections. A brief discussion of the background on related algorithms will be presented in Sect. 2. Section 3 will describe our new algorithm that consists of three parts, namely, identifying the compromised nodes in a targeted WSN and the location of the compromised node with the known beacons, then with the defining the resiliency degree it is to control the targeted WSN. Section 4 describes the simulation results and how to control the resiliency of WSN by the required parameter that can be completed by disabling the compromised nodes. A briefly discussion is given with our conclusion in Sect. 5.

2 Internal Attacks in WSN

Simple sensor nodes are usually not well physically protected due to they are cheap and are always deployed in open or even in hostile environments where they can be easily captured and compromised, hence, an adversary can extract sensitive information, control the compromised nodes and let those nodes service for the attackers. The attacks are involved in corrupting network data or even disconnecting major part of the network.

Karlof and Wagner discussed attacks at the network layer in [11] and mentioned altered or replayed routing information and selective forwarding, node replication, Sybil attacks or black-grey-sink holes, and HELLO flooding. Some papers discussed various attacks in term of network's resiliency, such as [12, 13], discussed how to keep WSN routing protocols as stateless as possible to avoid the proliferation of specific attacks and provide for a degree of random behavior to prevent the adversary from determining which the best nodes to compromise are.

Unlike traditional routing, where intermediate nodes just forward input packets, in network coding intermediate nodes actively mix or code input packets and forward the resulting coded packets. The very nature of packet mixing also subjects network coding systems to a severe security threat, knows as a pollution attack, where attackers inject corrupted packets into the network. Since intermediate nodes forward packets coded from their received packets, as long as least one of the input packets is corrupted, all output packets forwarded by the node will be corrupted. This will further affect other nodes and result in the epidemic propagation of the attack in the network. In [12], it addressed pollution attacks against network coding systems in wireless mesh networks.

There are a few papers also addressed pollution attacks in internal flow coding systems use special crafted digital signatures or hash functions. Recently some papers discuss the preventing the internal attacks by related protocols.

It is noted that resiliency of WSNs are related to the security of WSNs, where a definition of network resiliency was discussed based on the comparisons with other similar terminologies such as robustness etc. In this paper a definition was presented, i.e. resiliency is the ability of a network to continue to operate in presence of k compromised nodes. We shall use this definition to show a definition of "resiliency degree" and then present an algorithm to handle the compromised nodes to ensure the targeted WSN working at the required level.

In terms of the attack model, synoptically speaking there are two types of international attacks, namely (a) exceptional message attack, by which the attacks will tamper the message content or generate fake messages and (b) abnormal behavior attacks, by which the transmission will be abnormally changed such as dropping the messages, forwarding the message to a particular receivers, broadcasting redundant or meaningless messages to increasing the traffic load in the network, etc. As we are focusing on the robust resiliency based on the internal attackers we shall focus on the case abnormal attributes and some of cases described in exceptional message attack can be extended to what we discussed in this paper.

3 Novel Algorithm for Protecting WSNs from the Internal Attacks

The system under consideration consists of an area of interest where region wise detection requirements are provided by the end user. We model the area of interest as a grid Ω of $N_x \times N_y$ points. The ratio of the detection to miss requirements at every point on the grid are ordered in two $N_xN_y \times 1$ vector of the ratio of the probability, $\mathbf{p}_\mathrm{d}/\mathbf{p}_\mathrm{m}$. There are two common sensing models found in literature, binary detection model and the exponential detection model. Both models share the assumption that the detection capability of a sensor depends on the distance between the sensor and the phenomena, or target to be detected. Following [12] notations we have the case that for the binary detection model, the probability of detection p_d (t,s) is given as:

$$p_d(t,s) = \begin{cases} 1 & \text{if } d(t,s) \le r_d \\ 0 & \text{if } d(t,s) > r_d \end{cases} \tag{1}$$

where r_d is the detection radius and $d(t,s)$ is the distance between the target's position "t" and the sensor location "s" on a plane. The exponential model is a more realistic model, where the probability of detection corresponds to

$$p_d(t,s) = \begin{cases} e^{-\alpha d(t,s)} & \text{if } d(t,s) \le r_d \\ 0 & \text{if } d(t,s) > r_d \end{cases} \tag{2}$$

where α is a decay parameter that is related to the quality of a sensor or the surrounding environment. In the exponential model of Eq. (2), even if a target is within the detection radius, there is a probability that it will not be detected, which means it will be missed. As this model is closer to the realistic case, we shall use this model.

The process of linking individual sensors' detection characteristic to the overall probability of detection requirements on the grid is mathematically quantified using miss probabilities, $p_\mathrm{miss} = 1 - p_\mathrm{d}$, where p_d is the probability of detection. The overall miss probability $M(x, y)$ corresponds to the probability that a target at point (x, y) will be missed by all sensors, which is

$$M(x,y) = \prod_{(i,j)\in\Omega} p_{miss}((x,y),(i,j))^{u(i,j)} \tag{3}$$

where $u(i, j)$ represents the presence or absence of a sensor at the location (i, j) on the grid, and corresponds to

$$u(i,j) = \begin{cases} 1, & \text{if} \quad \text{there is a sensor at } \ (i,j) \\ 0, & \text{if} \quad \text{there is no sensor at } \ (i,j) \end{cases} \tag{4}$$

Taking the natural logarithm of the both sides in Eq. (3), we have

$$m(x, y) = \sum_{(i,j)\in\Omega} u(i, j)\ln p_{miss}((x, y), (i, j)) \tag{5}$$

where $m(x, y)$ is so-called the overall logarithmic miss probability at the point (x, y). Thus we have the function $b(x, y)$ as

$$b(x, y) = \begin{cases} \ln p_{miss}((x, y), (0, 0), d((x, y), (0, 0) \leq r_d \\ 0, d((x, y), (0, 0)) > r_d \end{cases} \tag{6}$$

The overall logarithmic miss probabilities for all points on the grid can be arranged in a vector **m** of dimension $N_x N_y \times 1$ that corresponds to Eq. (7) as shown below:

$$\mathbf{m} = [m(x, y), \forall(x, y) \in \Omega]^T$$
$$\mathbf{u} = [u(i, j), \forall(i, j) \in \Omega]^T$$

and

$$\mathbf{m} = \mathbf{Bu} \tag{7}$$

The $((i - 1)\ N_y + j)$th element of **u** indicates the number of sensors deployed at point (i, j) on the grid. The matrix **B** is of dimension $N_x N_y \times N_x N_y$, and it contains

$$\{b(x - i, y - j), \forall(x, y) \in \Omega, (i, j) \in \Omega\}$$

$b(x - i, y - j)$ corresponds to the (r, c)-th entry of **B**, where $r = (x - 1)\ N_y + y$ and $c = (i - 1)N_y + j$.

Essentially, $b(x - i, y - j)$ quantifies the effect of placing a sensor at the point (i, j) on the logarithmic miss probability at the point (x, y) on the grid. If there are some compromised nodes distributed in a WSN, how those compromised nodes could be detected by their so-called abnormal attributes among the network, such as irregular change of hop count that implicates sinkhole attacks; the signal power is impractically increasing which may indicate wormhole attacks; abnormally dropping rate traffic behaviors related the related nodes most likely to be compromised, etc. First we propose each sensor node can establish pair wise keys with its one-hop neighbors. When the nodes are deployed, each node is pre-distributed an initial key, K_I. A node, Q_i with $Q_i \in \Omega$ can use K_I and one way hash function H_f to generate its master key, K_M:

$$K_M = H_f(ID_i)K_I \tag{8}$$

Here we highlight the identification of Q_i is ID_i in the above equation. Then node Q_i broadcasts an advertisement message (ID_i, $Nonce_i$) which contains a nonce, and waits for other neighbor Q_j (here $i \neq j$) to respond with its identity. So the process will be as follows:

$$
\begin{aligned}
Q_i &\Rightarrow * : ID_i, Nonce_i \\
Q_j &\Rightarrow Q_i : ID_{j,} MAC(K_j, ID_j | Nonce_i)
\end{aligned}
$$

Therefore, at the same time, Q_j can also generate the key K_j. Then both nodes Q_i and Q_j can generate the pair-wise key $K_\mathrm{i,\,j} = H_\mathrm{f} K_\mathrm{j}(ID_\mathrm{i})$. So each node can use these nodes' ID to calculate its one hop-neighbors' key, i.e. $\forall Q_\mathrm{i} \in N_1$, where N_1 is the space of one-hop for a fixed node in the targeted SWN. If there is any stranger node, such as the adversaries' node, it will be distinguished by those pair-wise keys.

In this case we need to check those nodes who are sitting in the group belong to more than one hop neighbor. There are many ways to do this for example in [12] there is a filter used via so called "minimum acceptable trust value." In order to reduce the computing costs, we believe that the second order parameter will be good enough to prevent form compromised nodes, where $\forall Q_\mathrm{i} \notin N_1$, as it will be ultimately checked by abnormal attributes. It is noted that a target node can only send data packets over the sensor channel. The neighboring sensor nodes, which are within the sensing radius of the target node as shown in Eq. (2), will then receive the stimulus over the sensor channel. That receives and detects over the sensor channel has to forward its sensing results to one or more sink nodes over the wireless channel. Inside a sensor node the coordination between the sensor protocol stack and the wireless protocol stack is done by the sensor application and transport layers. Any networking processing mechanism can be implemented in the sensor application layer.

Assume that the $S_\mathrm{i}/N_\mathrm{i} = R_\mathrm{i}$ for a reference node $Q_\mathrm{i} \in N - N_1$, when $N \subset N_\mathrm{multi\text{-}hop}$ and $N_1 \subset N$. we have the following parameters:

$$\overline{\mu}_j = \frac{1}{n_1} \sum_{i=1}^{n_1} R_i \tag{9}$$

$$\overline{\sigma}_j = \sqrt{\frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (R_i - \overline{\mu}_j)^2} \tag{10}$$

Then we have

$$CK_j = \left| \frac{R_i - \overline{\mu}_j}{\overline{\sigma}_j} \right| \tag{11}$$

If CK_j is smaller than the designed threshold, it would be taken as normal case otherwise it would be assume the checked node $Q_\mathrm{i} \notin N_1$ and $Q_\mathrm{i} \in N$. We defined the transmission rate, T_i as the *i*th node in the targeted SWN to express its transmission attribute:

$$T_i = \frac{T_i^{out}}{T_i^{In}} \tag{12}$$

Here, the up script "out" and "In" denoted the signal sending to next one-hop node and received from pervious node. It is noted that equation does not loss the generality for example, in the case that multi-one hop nodes to be sent, the transmission will follow the designed particular protocol, where is beyond the scope of the current paper. The checking threshold will take the consideration for the designed protocol. So we can let the signal "out" and "In" situation absorbed by the particular protocol. Then we can simply apply the Eqs. (9)–(11) subject to R_i replaced by T_i. For example, it is noted that if T_i is out of the mean value the value obtained by Eq. (10), it will be taken by compromised node if there is no pervious knowledge about the SWN. Also if we have the information about the SWN empirical data we may use the Chi-square curve to check the fitting.

Following the definition of resiliency is the ability of a network to continue to operate in presence of k compromised nodes; we assumed the threshold for our currently targeted SWN is designed as 30 % of the total nodes became compromised nodes. We may describe those sick nodes become a group denoted as S_q, where q is the qth sick group. The operation for resilient SWN is

$$\forall Q_i \in \sum_q S_q \geq 0.3N \tag{13}$$

With this condition, there is an operation needed to disable or isolated those compromised nodes by their locations in the targeted SWN.

Following our previous paper [13], we know that a WSN, for example, CC2431 includes hardware 'Location Engine' that can calculate the nodes position given the RSSI (Radio Signal Strength Indication) and position data of reference nodes. As described in [13] that another factor that affects received signal strength is antenna polarization, which is one of focuses of this paper. Small simple antenna's produce linearly polarized radiation. For the linear polarized case, the electrical magnetic (EM) field remains in the same plane as the axis of the antenna. When the antenna produces an electric filed that remains in a plane parallel with respect to earth the polarization is said to be horizontal.

4 Simulation and Discussion

We consider a homogenous WSN with 1024 sensors uniformly distributed in the network area, which is in the network region b by b squared field located in the normalized resiliency-degrees against the normalized time units. In order to investigate the interference effects to the WSNs, we take two cases, namely, 32 × 32 (low density case) and 16 × 16 (high density case) squared fields. The simulations were running 50 times with the final averaging the data as shown in Fig. 1, which is the case that "normalized average delivery rate" versus "percentage compromised nodes."

It is noted that the "series1" in the Fig. 1 is the chart about the average forward rate ≅55 % and the "serious 2" is the case average forward rate ≅32 %. At the same compromised node rate the latter case will be more serious than the former. There are two charts the "series 1" is the sensors deployed in the smaller area (16 × 16) and

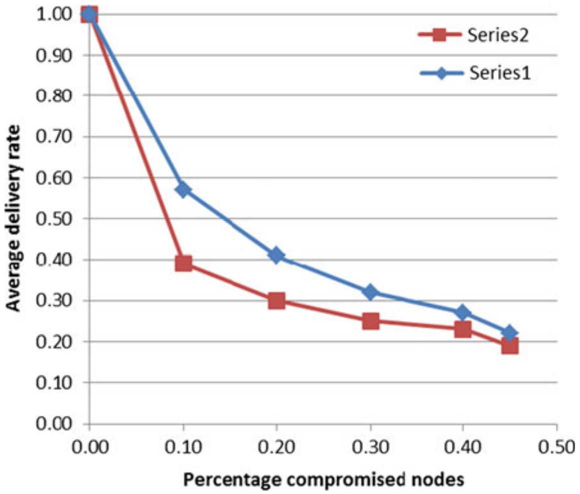


Fig. 1 Chart of "the "normalized average delivery rate" versus "percentage compromised nodes"

the same sensors were distributed in the larger area (32 × 32) is the case of "series 2". Due to the crowd sensors will impact each other by the interferences so the detection accuracy are impacted.

Figure 2 shows the situations about "normalized resiliency ration" versus "the simulation period time.

In our experiments, if it decreased by 30 % the operation is taken to identify the compromised nodes and disable them with their locations. In our experiments we have divided the whole areas by four regions each region we design three beacons (locations known) by which with RSSI to get the locations for compromised nodes and then disable them when the "operation is running." From the Fig. 2 we can see that the resiliency is under a reasonable level to be controllable.

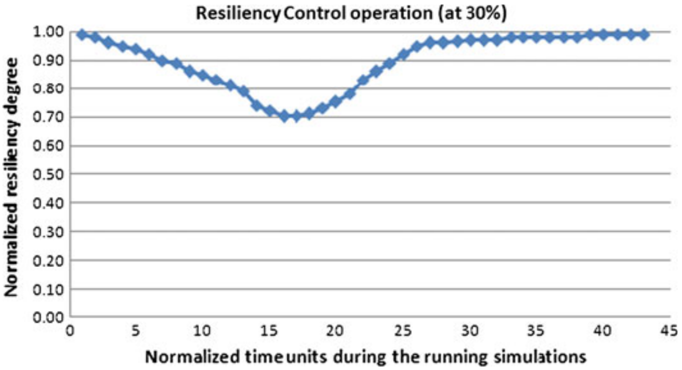


Fig. 2 Normalized resiliency degree

5 Conclusion

In this paper we have been focusing on controllable resiliency in a WSN, in particularly the internal attacks such as compromised nodes. We first discuss the method to identify those compromised nodes by abnormal attributes. Then follow our previous paper we used RSSI to find the locations for the conformed compromised nodes. We defined resiliency degree for a WSN based on the definition of paper [12]. Then for the WSN approaching the designed resiliency degree for the WSN, the operation is taken and the resiliency degree in the targeted WSN will be recovered as shown by Fig. 2.

References

1. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) Wireless sensor networks: a survey. Comput Netw 38(4):393–422
2. Chung-Kuo C, Overhage JM, Huang J (2005) An application of sensor networks for syndromic surveillance. Proc of Networking, Sensing and Control, pp 191–196
3. Werner-Allen G, Lorincz K, Ruiz M, Marcillo O, Johnson J, Lees J, Welsh M (2006) Deploying a wireless sensor network on an active volcano. IEEE Internet Comput 10:18–25
4. Sinopoli B, Sharp C, Schenato L, Schaffert S, Sastry SS (2003) Distributed control applications within sensor networks. Proc IEEE 91:1235–1246
5. Sikka P, Corke P, Valencia P, Crossman C, Swain D, Bishop-Hurley G (2006) Wireless ad hoc sensor and actuator networks on the farm, Proc of IPSN 2006, pp 492–499
6. Stephens DL Jr, Peurrung AJ (2004) Detection of moving radioactive sources using sensor networks. IEEE Trans Nuclear Sci 51:2273–2278
7. Feng Z (2004) Wireless sensor networks: a new computing platform for tomorrow's Internet. vol 1, pp 1–27
8. Huang X, Wijesekera S, Sharma D (2009) Fuzzy dynamic switching in quantum key distribution for Wi-Fi networks. In: Proceeding of 6th international conference on Fuzzy systems and knowledge discovery, Tianjin, China, pp 302–306, 14–16 Aug 2009
9. Huang X, Shah PG, Sharma D (2010) Multi-agent system protecting from attacking with elliptic curve cryptography. In: The 2nd international symposium on intelligent decision technologies, Baltimore, USA, 28–30 July 2010
10. Huang X, Sharma D (2010) Fuzzy controller for a dynamic window in elliptic curve cryptography wireless networks for scalar multipication APCC 2010. In: The 16th Asia-Pacific conference on communications, Langham Hotel, Auckland, New Zealand, pp 509–514, 31 Oct 3, Nov 2010. ISBN 978-1-4244-8127-9
11. Karlof C, Wagner D (2003) Secure routing inn wireless sensor networks: attacks and countermeasures. Ad Hoc Netw 1(2–3):293–315
12. Ochirkhand Erdene-Ochir, Marine Mibier (2010) Fabrice Valois and Apostolos Kountouris, Resiliency of wireless sensor networks: definitions and analyses. In: 17th international conference on, telecommunications. pp 828–835
13. Huang X, Barralet M, Sharma D (2009) Accuracy of location identification with antenna polarization on RSSI. In: International multi-conference of engineers and computer sciences 2009, IMECS 2009, Hong Kong, p 542, 18–19 March 2009

Channel-Usage Model in Underlay Cognitive Radio Networks

Sanjib K. Deka and Nityanada Sarma

Abstract In underlay cognitive radio networks (CRNs), a licensed channel can be shared with its primary users (PUs), provided the resultant interference power level at the primary receiver is below a predefined threshold, known as interference power constraint (IPC). Therefore, in such a scenario secondary users (SUs) have to perform the following—(a) perform spectrum sensing to detect the existence of PUs and (b) estimate the IPC to protect the PU from harmful interference. In this paper we develop a channel-usage model of PU's in underlay mode of CRNs using a Hidden Markov Model (HMM) that allows SU to predict the unused/usable spectra. The model estimates the interference power level of a license channel due to the presence of SUs and decides on the availability of the channel for SU's transmission imposing IPC. Further, a channel selection scheme is developed so that SUs can decide and select the best channel in the presence of multiple available licensed channels. Simulation results demonstrate the efficacy of the proposed model and its usability in underlay CRN.

1 Introduction

In Cognitive Radio Network (CRN), SUs along with the role to sense and discover spectrum opportunities, cognitive radio (CR) [1] has to detect the existence or return of the PUs to protect the PUs from potential interference. The policies for MAC-layer sensing [2] of SUs in underlay CRN require the information about the availability of licensed channels and the estimation of interference level. In such a scenario, the

S. K. Deka (✉) · N. Sarma
Department of Computer Science and Engineering, Tezpur University, Tezpur, Assam 784 028, India
e-mail: sdeka@tezu.ernet.in

N. Sarma (✉)
e-mail: nitya@tezu.ernet.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_12,

SUs require to learn about the channel's usage pattern of its PUs which will help them in predicting future channel availability, which helps to alleviate the sensing overhead problem of proactive sensing used by SUs.

In this paper, we focus mainly on two important issues of MAC-layer sensing for underlay mode CRN—(i) how to estimate and model the channel usage pattern of PUs while tolerating interference from SUs, and (ii) how to discover opportunities in the licensed channel based on learnt channel usage pattern. Accordingly, we propose a novel channel-usage model of PUs for use by the secondary nodes to discover the spectrum opportunity using Hidden Markov Model (HMM) [3]. The model addresses the issue of interference in a license channel due to the presence of SUs through estimated IPC [4, 5] to protect PUs. Based on the proposed model, a channel selection scheme is developed, which helps to decide and select the best channel to be used by a SU for transmission in future.

As shown in Fig. 1, for a spectrum underlay scenario, IPCs can be imposed to protect the PUs from interference. IPC [4, 5] is defined as the maximum interference power level tolerated by a PU-Rx (say the kth PU) which can be calculated as $Z_k = BT_k$, where, B is Boltzmann's constant, and T_k is a predefined threshold for the tolerable interference power by the PU-Rx. Now, let $g_{i,k} = E\{|g_{i,k}|^2\}$ be the channel gain for the link between the ith SU-Tx to the kth PU-Rx ($E\{.\}$ denote statistical expectation) and $p_i = E\{|x_{s,i}|^2\}$, (where, $x_{s,i}$ denote the transmitted signal from SU i) be the transmit power of the SU i. Using the definition of IPC, the total interference power received by any PU-Rx from reaching a harmful level, the IPC, I_t ($I_t = \sum_{i=1}^{N} p_i g_{i,k} \leq Z_k, k = 1, \ldots, M_k$) must be satisfied by the SUs [6]. After presenting system overview and the assumptions for the model, the rest of the paper is organized as follows. In Sect. 3 we present the proposed channel-usage model of PUs. The channel selection scheme and validation of the model is presented in Sect. 4. Section 5 discusses the related work and conclusion of the paper with some future directions.

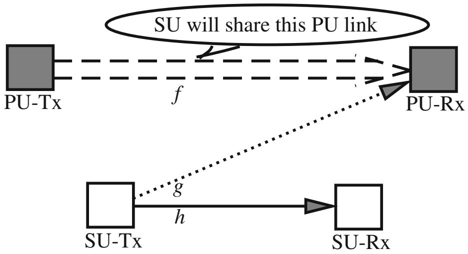


Fig. 1 Spectrum sharing between PU and SU. Links consisting of PU transmitter (PU-Tx) and receiver (PU-Rx), SU transmitter (SU-Tx) and receiver (SU-Rx) and SU-Tx and PU-Rx are represented with channel gains f, h and g respectively

2 System Overview

A single-hop wireless network is considered as the secondary network (SN) with a group of SUs equipped with sensor and single tunable antenna (with range to cover N licensed channels). It is also assumed that within the transmission range of the SN there are no other SNs interfering or cooperating with it. We consider that a secondary node N_0 has M neighbors $(N_1, N_2, \ldots, N_M)$, where any one of N licensed channels can be used by N_0 depending on the availability at that particular point of time and perform asynchronous transmission. A data link L_i $(i = 1, \ldots, M)$ is used for communication between N_0 and N_i. We consider a basic collaboration policy that let SUs to participate in sensing a channel overcoming the situation of channel fading. We assume that the estimation of IPC by SUs use the average interface power (AIP) techniques of IPC computation to tackle the fading states of the links. We assume that the SUs are equipped with power control scheme with tunable capability that can adjust the power inflict on PUs. In practice, the channel gain g can be obtained at SU-Tx through an estimation of the received signal power from PU-Rx when PU-Rx transmits, under the assumptions of the pre-knowledge of the PU-Rx transmit power level and channel reciprocity [5].

3 The Proposed Model

A channel can be represented by a string of 0s and 1s, which are the symbols observed by a SU. The length of the string is determined by the length of the sensing period. Symbols 0 and 1 represent the usable portion and busy portion of the channel respectively.

A SU scans and senses N licensed channels periodically and computes the IPC [4, 5] for each of the channel. It then compares the computed IPC with the predefined threshold at the primary receiver and constructs the string for a channel using Eq. (1). Assuming that for a licensed channel C to kth PU-Rx, at any given observation instance t, O_t be the observation symbol seen by a SU. The channel dynamics using the IPC can be defined as follows:

$$O_t = \begin{cases} 0 & \text{if } I_t \leq Z_k \\ 1 & \text{if } I_t > Z_k \end{cases} \tag{1}$$

The secondary node records the licensed channel observation according to Eq. (1). The whole observation time is divided into equal sized periods, and each of the period (i.e. sensing period, say, T_P^c) in turn is divided into T number of observation slots as shown in Fig. 2. The observation in a given slot is represented by 0 or 1. Therefore, over a period of T time slots, we obtain the observation sequence O, as $O = O_1\, O_2 \ldots O_T$. where, $O_t \in [0, 1], \forall t = 1 \ldots T, T =$ number of time slots used for observation. It will generate a string of 0s and 1s representing the sensed channel.

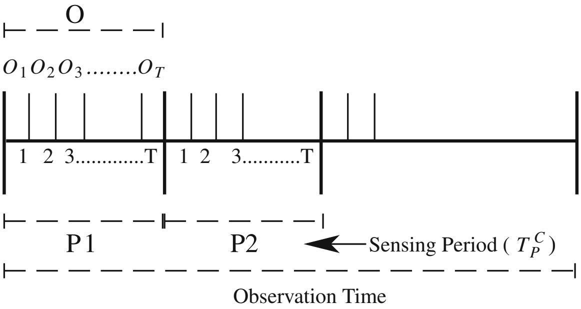


Fig. 2 The licensed channel observation sequence

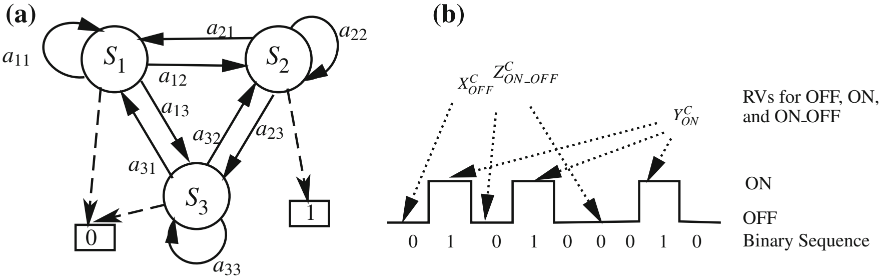


Fig. 3 **a** A licensed channel as an HMM. **b** Illustration of sensing a channel

Therefore, for the first sensing period, O can be represented as $O = \{O_t\}_{t=1}^{T}$. Now, for the next period, we can derive the observation sequence as, $O_{next} = \{O_t\}_{t=T+1}^{2T}$. More the number of 1s that would be recorded in an observation sequence, the higher would be the interference indicated to a PU by a SU. The recording of the observation follows the behavior of a Markov chain and can be represented by a hidden Markov chain.

For an underlay CRN, the PU can be at any one of the three states at a time as the channel usage pattern followed by PUs can be of three fold. A PU can be active on a channel without allowing the possibility of secondary access or with the permissible secondary access. The other possibility is that the PU can be completely off on the channel during the SU sensing period. Therefore, a channel used by its PUs can be modeled as a Markov process represented by an ergodic Hidden Markov Model (HMM) [3] with state space of three. As shown in Fig. 3a, the PU can be at any of OFF, ON, and ON_OFF (i.e., partially ON/OFF state determined by IPC tolerable to the PU) states for some instance t on a channel and are represented by states S1, S2 and S3 of a HMM. For a channel $C(C = 1, 2, \dots N)$, we can model the sojourn time of OFF, ON and ON_OFF periods as random variables and are decided by the associated IPC (i.e. I_t) values during then. The distribution of IPC values can also be modeled as a random variable. Thus the joint probability density functions (p.d.f.) of respective random variables and IPC represent the duration OFF, ON and ON_OFF periods. When the PU moves from one state to another, the transition between states obeys by conditional probabilities based on the values of the IPC during then and decided the sojourn time of a PU that remains in a state.

Let, H_c be the HMM for a channel C. It can be represented as, (A, B, π), where, N is the number of states in the model (3 in our model). M is the total number of distinct observation symbols (two symbols, 0 and 1 in our model). A is the state transition probability matrix with probability values given by conditional p.d.f. of random variables of S_1, S_2 and S_3. B is the observation symbol probability distribution matrix, π is the initial state probability distribution vector with probability values given by p.d.f.s of random variables of S_1, S_2 and S_3.

According to the principle of HMM, H_c is composed of two stochastic processes, possibilities of producing state sequences and possibilities of producing observation symbol sequences constituting the string of 0 and 1s (observable sequences). However, the strings of 0 and 1s represent the channels under consideration and is required to be optimized for consideration. The optimization can be done by maximizing probability of observation symbols produced by H_c, which eventually maximizes the HMM. This has lead to the estimation of the HMM parameters A, B and π as follows.

To estimate state transition probability matrix, A, let the random variables representing the distribution of S1, S2 and S3 of a channel C and IPC are X^c_{off}, Y^c_{on}, $Z^c_{on\_off}$ and I respectively. So, the random vectors (absolute continuous random vector) for the states S_1, S_2 and S_3 are $[r, I]$, where, $r \in [X^c_{off}, Y^c_{on}, Z^c_{on\_off}]$, $X^c_{\mathit{off}}, Y^c_{on}, Z^c_{on\_off} \in T^c_P$, and $I \in I_t$. Mathematically, $A = [a_{ij}], 0 < i, j \leq N$ with $a_{ij} = P(S_i|S_j), 0 \leq a_{i,j} \leq 1$ and $\sum_{j=1}^{N} a_{ij} = 1$. For a given, $I = i$, the probabilities of A can be obtained with the conditional p.d.f.s of random variables X^c_{off}, Y^c_{on}, and $Z^c_{on\_off}$ respectively. Then the probabilities can be $P(r \in T^c_P|I = i)$. Let, the joint p.d.f. of $r, I \in [r, I]$ be $f_{rI}(r, i)$ and the marginal p.d.f. of $I \in I_t$ be $f_I(i)$. The conditional p.d.f.s could be derived such that,

$$f_{r|I=i}(r) = \begin{cases} \frac{f_{rI}(r,i)}{f_I(i)}, & \text{if } f_I(i) > 0 \\ 0, & \text{Otherwise.} \end{cases} \tag{2}$$

So, depending on the value of i compared to Z_k the IPC condition is maintained and the Eq. (2) will decide the state transition probabilities.

Let $v_t \in \{0, 1\}$ be the observation symbol. The probabilities (observation probabilities) of producing these symbols depend on the p.d.f. of random variable I provided it's value is compared to the Z_k. Mathematically, $B = \left[b_j(k)\right], 0 < j \leq N$, $0 < k \leq M$ with $b_j(k) = P(v_t|S_j), 0 \leq b_j(k) \leq 1$ and $\sum_{k=1}^{M} b_j(k) = 1$.

The initial probability distribution vector, π, can be derived from the p.d.f. of the random variables respective to the states of the HMM. Therefore, π can be obtained as, $\pi = \{\pi_j\}$, where, $\sum \pi_j = 1, 0 < j \leq N$. So, $\pi = \{P(X^c_{\mathit{off}}), P(Y^c_{on}), P(Z^c_{on\_off})\}$.

Estimating the parameters, the HMM can be trained with recorded observation sequence, O using the Baum-Welch [3] estimation procedure to get Expectation-

Maximization (EM) of the observation. Baum-Welch procedure will compute the log probability of observation by the estimated model to get the maximum value. The trained HMM is validated for stability and accuracy monitoring the log-likelihood of occurrence of each unseen test sequences. Again, it can be checked that the log-likelihood values for the test sequences would be higher when computed using trained model in comparison to untrained model. Once a credible trained HMM is determined, it can be used to sense a channel to check its availability for SUs. Figure 3b, illustrates the periodic sensing of a channel using the model.

4 Channel Selection Scheme and Validation of the Model

An availability metric (AM) for a channel C, say AM_C, can be defined in such a way that the lesser the number of symbol 1 present in a generated channel sequence, the lower the interference indicated for the channel at the SU and hence the channel be available for SU. Further, usability of a channel is also determined from the separation between occurrence of two symbol 0s. AM of a channel C can be computed using the formula, $AM_C = U_{seq} + \frac{L_{seq}}{S^0_{seq}}$, where, U_{seq} = average separation between two 0s in a sequence string, L_{seq} = length of the sequence string (i.e., the number of time slots, T, used in the sensing period, T^c_P), S^0_{seq} = number of symbol 0 in a sequence string. For example, if a channel is represented by a 8-bit string say 01001101, then $U_{seq} = (1 + 0 + 2)/3$ and eventually $AM_c = 1.5$. Therefore, the channel selection scheme can be framed with the following steps. First, obtain a credible trained HMM for each licensed channel C. Second, generate sequences by the HMM and predict for the channels. Third, compute AM for the generated sequences using the above formula. Fourth, for each sequence, if the AM value for a channel C is the smallest (means less interference) amongst all the channels, the channel C would be selected by SU as the preferred channel.

While an SU has data to transmit to one of its neighbor, it senses the licensed channel on-demand having the smallest AM value, consulting with the model. At every stage, the model generates sequences and predicts the availability of the channels for the next time slot using their AM values. If the channel is found to be free the SU will continue transmission. Otherwise, it will use the model to find the next channel to be sensed physically.

To validate the proposed model, we simulate a single hop wireless sensor network with 5 nodes in *ns2* [7] simulator. The nodes are configured to move randomly and to generate traffic in random using the *s-mac* [8] protocol. We simulate two licensed channels at frequency of 900 MHz, bandwidth 200 KHz, and supports data rate of 270 Kbps. All the 5 nodes are configured to transmit on the channels. One of the 5 nodes is assumed as a SU at a time and observes the other four's activity on a channel and eventually records the channel observation sequences using the symbols 0 and 1. This is done using the RTS-CTS frames of *s-mac* protocol. Similarly, we record the training and test sequences for the two licensed channels, say $ch1$ and $ch2$ and are used to validate our proposed model. For our experiment, we represent the channels with 296-bit and 307-bit long binary strings and hence the length of training and test sequences are 296 and 307 symbols respectively. Hence, we take, $T = 296$ and 307. We also assumed that the interference power threshold value for our experimental region to be 10^{-8} W. For the experiment, $ch1$ and $ch2$ is represented with HMMs setting with initial probability values and are trained. From the experiments, it is observed that Baum-Welch algorithm takes 18 and 10 numbers of iterations to maximize the log-probability (Expectation Maximization) with training sequences of $ch1$ and $ch2$. We have used 15 recorded test sequences (i.e., previously unseen sequences) per channel to validate the trained HMMs for the channels. Figure 4, shows the plot of log-likelihood of the test sequences for the channels $ch1$ and $ch2$ and its variation that become minimal starting from close to their 5th sequences. This results show that the trained HMMs can be considered statistically stable and accurate which proves that the HMMs have been trained properly. As long as the same IPC conditions prevail for both the channels, using more and more number of test sequences the stability of the HMMs can be proved. Similarly, we can represent other possible channels with HMMs and use them for future channel availability predictions. Figure 5 demonstrates the availability metric values for channel sequences generated by the trained HMMs for both the channels $ch1$ and $ch2$. $ch2$ has smaller AM values for channel sequence numbers 1, 2, 3, 6, 7, and 9 compared to $ch1$. Therefore, $ch2$ will be the preferred channel by the SU for transmission in those future time periods. Similarly, the SU will prefer $ch1$ over $ch2$ for transmission in the future time periods 4, 5, 8 and 10. These results demonstrate the applicability of the proposed channel usage model to predict for licensed channels by SUs within tolerable interference to PUs.

5 Related Work and Conclusion

Most of the existing works [9–16] propose mechanisms to discover unused spectrum holes in licensed band using the overlay mode access and considers ON-OFF channel usage model. Authors of [17] consider the underlay mode discovery of opportunity

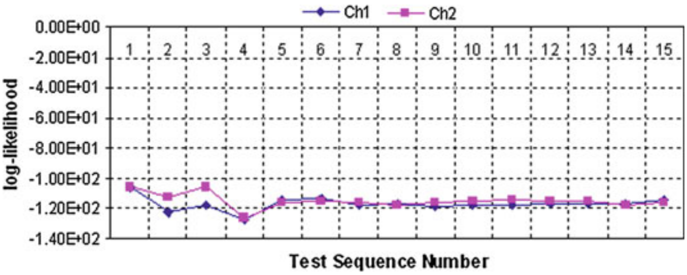


Fig. 4 Log-likelihood (ℓ_i) for test sequences for channels $ch1$ and $ch2$

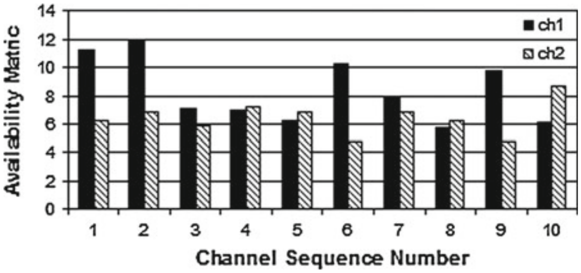


Fig. 5 Availability metric for channels *ch*1 and *ch*2 with different IPC conditions, showing different AM values for different channel sequences

in licensed bands, but they do not recognize the interference and noise into the channel. Authors of [18] propose a technique to model a channel with consideration to interference using Hidden Markov Model. However, they did not consider a practical channel model that would maximize the discovery of opportunities during on demand spectrum sensing.

In this paper, we have presented an HMM based channel-usage model for a licensed channel under the interference power constraint to discover opportunity by secondary (unlicensed) users in underlay CRN. To study the effectiveness, we have trained and validated the model with simulations. The experimental results show that the proposed model can be used by a secondary node for MAC layer sensing to decide the sensing order to locate an available channel with minimum delay. Sophisticated collaboration mechanism among SUs are necessary for enabling dynamic channel switching, which is left as part of future work. Further, performance evaluation of the model in a real testbed would be more interesting.

References

1. Mitola J, Maguire G (1999) Cognitive radio: making software radios more personal. IEEE Pers Commun 6(4):13
2. Deka SK, Sarma N (eds) (2011) A survey on MAC protocols for cognitive radio networks. National Conference on Trends in Machine Intelligence (NCTMI), Assam, India
3. Rabiner LR (ed) (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 77(2):257–286
4. Tian Z, Yang C-G, Li JD (2010) Optimal power control for cognitive radio networks under coupled interference constraints: A cooperative game-theoretic perspective. IEEE Trans Veh Technol 59(4):1696
5. Zhang R (2009) Cooperative feedback for multiantenna cognitive radio networks. IEEE Trans Wirel Commun 8(4):2112
6. Zhang Z, Chen HH, Chen Y, Yu GG, Qui PL (2008) On cognitive radio networks with opportunistic power control strategies in fading channels. IEEE Trans Wirel Commun 7(7):2752
7. ns2. http://www.isi.edu/nsnam/ns/
8. Heidemann J, Ye W, Estrin D (2004) An energy-efficient mac protocol for wireless sensor networks. IEEE/ACM Trans Netw 12(3):493

9. Yang Z, Hamid M, Mohammed A (ed) (2010) On spectrum sharing and dynamic spectrum allocation : MAC layer spectrum sensing in cognitive radio networks In: International conference on communications and mobile, computing, 2010
10. Kim H, Shin KG (2008) Efficient discovery of spectrum opportunities with MAC-layer sensing in cognitive radio networks. IEEE Trans Mobile Comput 7(5):533
11. Fu H, Li H (ed) (2009) An adaptive sensing period algorithm in cognitive radio networks. In: Proceedings of ICCTA2009, IEEE, 2009
12. Kim H, Shin KG (2006) Adaptive mac-layer sensing of spectrum availability in cognitive radio networks, department of electrical engineering and computer science, university of michigan, ann arbor, mi. Tech. Rep. CSE-TR-518-06, Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA (2006)
13. Tranter W, Akbar IA (2007) Dynamic spectrum allocation in cognitive radio using hidden markov models: Poisson distributed case. In: Proceedings of IEEE Southeast conference, pp 196–201
14. S.M.L.M.S.S. Chang-Hyun Park, Sang-Won Kim, Microwave Conference, APMC 2007. Asia-Pacific pp 1–4 (2007)
15. Hattangadi SM, Liu J, Wei Y (2011) In: ICCCTA. (http://www.ir.bbn.com/projects/xmac/rfc/rfcaf.pdf, 2011)
16. Wang B, Ji Z, Liu K, Clancy T (2009) Primary-prioritized Markov approach for efficient and fair dynamic spectrum allocation. IEEE Trans Wirel Commun 8(4):1854. doi:10.1109/TWC.2008.080031
17. Pla V, Vidal JR, Martinez-Bauset J, Guijarro L (2010) Modeling and characterization of spectrum white spaces for underlay cognitive radio networks. In: Proceedings of IEEE international conference on communications. ICC 2010, pp 1–5. doi:10.1109/ICC.2010.5501788
18. Manuj Sharma AS, Nayak KD (eds) (2008) Channel modeling based on interference temperature in underlay cognitive wireless networks. In: Proceedings of IEEE international symposium on ISWCS (2008)

Supporting LTE Networks in Heterogeneous Environment Using the Y-Comm Framework

Mahdi Aiash, Glenford Mapp, Aboubaker Lasebae and Ameer Al-Nemrat

Abstract There are two trends in the research to develop future networks. While the first aims to introduce new technologies such as the Long Term Evolution (LTE) and WiMAX with high-speed data. The second aimed at providing clients with a ubiquitous connectivity via proposing new communication architectures to integrate different networking technologies and enabling mobile devices to switch seamlessly between them. Examples of such architectures are Y-Comm, Mobile Ethernet and IEEE 802.21. In this paper we will show how these research trends could be integrated. This is achieved by discussing how future communication frameworks like Y-Comm could fulfil the requirements and provide the functionalities of newly introduced technologies such as UMTS and LTE networks.

1 Introduction

In order to enhance user experience in future and Next Generation Networks (NGNs), a large number of research groups have been working on developing new communication mechanisms. The research effort in this field was divided into two directions. In the first, the research was concerned with enhancing the capabilities of current

M. Aiash (✉) · G. Mapp · A. Lasebae
School of Engineering and Information Science, Middlesex University,
Hendon, London
e-mail: M.Aiash@mdx.ac.uk

G. Mapp
e-mail: G.Mapp@mdx.ac.uk

A. Lasebae
e-mail: A.Lasebae@mdx.ac.uk

A. Al-Nemrat
School of Architecture, Computing and Engineering, University of East London,Newham, London
e-mail: a.al-nemrat@uel.ac.uk

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_13,

technologies and developing new ones with a high speed and low latency. This led to the evolution track of 2G, 3G and finally 4G networks such as the LTE [1]. While the second followed rather a different approach. Thus, in stead of developing a new brad technology or addressing the shortages of current ones, the research effort concentrated on introducing novel communication frameworks to integrate different wireless technologies and enabling future mobile terminals, which will be multi-homed by nature [2], to seamlessly switch between these technologies using vertical handover mechanisms. As a result of this research effort, novel communication architectures for Next Generation Networks (NGNs) have been proposed such as Y-Comm [13], the Ambient Networks [3]; the Mobile Ethernet [4] and IEEE802.21 [5].

In order to support the integration of a wide variety of wireless networks and enable the communication in such heterogeneous networks, each of the newly proposed communication frameworks has introduced a generic network architecture which provides the required functionalities to accommodate different wireless technologies and to support the mobility between them. However, as stated in [6], Y-Comm is the most completed one with a well-structured architecture that supports the provision of security, QoS, mobility and communication in an integrated manner. Therefore, this paper will consider Y-Comm as a representative of communication frameworks in heterogeneous networks. To the best of our knowledge, this is the first research work to bring the two research trends together and discuss how future communication architectures, the Y-Comm in this paper, could support newly introduced technologies such as the LTE.

The rest of this paper is organized as follows: Sect. 2 describes the LTE network infrastructure and its networking entities. Section 3 gives an overview of the Y-Comm framework and describes the Y-Comm's network structure along with its operational network entities. In order to show how Y-Comm could accommodate the LTE, Sect. 4 maps the Y-Comm framework onto the LTE Infrastructure and highlights the correspondence between the operational network entities in the LTE and Y-Comm network infrastructures. The paper concludes in Sect. 5.

2 The Long Term Evolution

2.1 An Overview

LTE is the preferred development path of GSM, Wideband Code Division Multiple Access (W-CDMA) and the High Speed Packet Access (HSPA) [7] networks currently deployed. This essential evolution will enable networks to offer the higher data throughput to mobile terminals needed in order to deliver new and advanced mobile broadband services.

LTE only supports packet-switched services and it aims to provide seamless Internet Protocol (IP) connectivity between user equipment (UE) and the packet data network (PDN), without any disruption to the end users' applications during mobility. The LTE networks resulted from the evolution of the Universal Mobile Telecommunications System (UMTS) radio access through the Evolved UTRAN (E-UTRAN) as well as the evolution of the non-radio aspects under the term System Architecture Evolution (SAE), which includes the Evolved Packet Core (EPC) network. Together LTE and SAE comprise the Evolved Packet System (EPS) [8].

2.2 Network Architecture

For both UMTS and LTE, the basic network architectures are very similar. In comparison to UMTS, the network elements used for LTE are upgraded and mostly renamed. However, they fulfil the analogous tasks in both cases [9]. The LTE network comprises three parts. Firstly, The EPC or the Core Network (CN) which contains several network elements such as the Home Subscriber Server (HSS), the PDN Gateway (P-GW), the Serving Gateways (S-GWs) and the Mobility Management Entity (MME). Each of these elements has different role which will be described later in this paper. Secondly, the Radio Access Network (RAN) or the E-UTRAN contains a number of base stations called eNodeBs (eNBs) and their controlling units which are directly connected to the Core network. Each eNB is connected to one or more MMEs/S-GWs with the S1 interface as shown in Fig. 1. Thirdly, the User's Equipment (UE) (also called the Mobile Terminal (MT)) includes the mobile device and a tamper-resistant card and the Universal Subscriber Identity Module (USIM) [10, 11]. Similar to the SIM card in 2G technology, the USIM is issued by the mobile operator and used to store security-related information that will be used to identify the subscriber.

2.2.1 The Elements of the Core Network

It is responsible for controlling the UE, establishing different bearers for different sessions while maintaining the desired QoS as well as authenticating subscribers to access the network resources. The CN includes the following logical elements [8]:

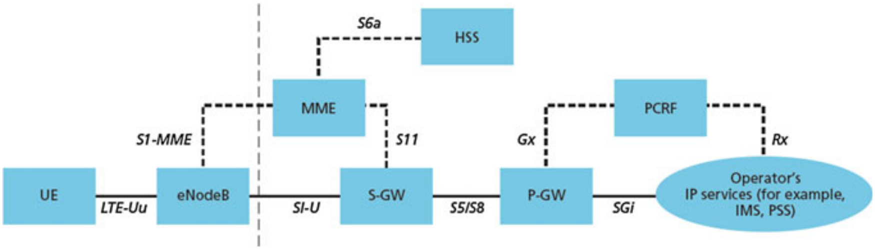


Fig. 1 LTE network architecture [8]

- **The Mobility Management Entity (MME)** manages session states, authentication, paging, mobility with 3GPP, 2G and 3G networks.
- **The Home Subscriber Server (HSS)** stores and manages all users' subscriber information such as QoS of different bearers, any roaming restriction. The HSS also contains dynamic information such as the identity of the MME to which the user is currently connected to.
- **The PDN Gateway (P-GW)** performs a per-user packets filtering and QoS enforcement for guaranteed bit rate (GBR) bearers, assigns IP address to UEs and serves as the mobility anchor for interworking with non-3GPP technologies. This implies that P-GWs are responsible for managing the so-called Vertical Handover which happens when a mobile device moves between different networking technologies such as LTE and WIMAX.
- **The Policy Control and Charging Rules Function (PCRF)** is a software component that operates at the P-GWs. It is responsible for policy control decision-making, as well as for controlling the flow-based charging functionalities.
- **The Serving Gateway (S-GW)** acts as a local Mobility Anchor point for inter-eNB handover and performs some administrative functions in the visited network such as collecting information for charging (for example, the volume of data sent to or received from the user) and lawful interception. The S-GW also acts as a mobility anchor for mobility within 3GPP technologies such as LTE and UMTS.

These elements are interconnected over a well-defined interfaces as shown in Fig. 1.

3 The Y-Comm Framework

The Y-Comm Framework is a communication architecture to support vertical handover for multi-homed nodes in heterogeneous environment. The architecture has two frameworks:

- The Peripheral framework deals with operations on the mobile terminal.
- The Core framework deals with functions in the core network to support different peripheral networks.

A brief explanation of Y-Comm is now attempted starting with the lowest layer. A more detailed explanation can be found in [12].

3.1 The Peripheral Framework

As shown in Fig. 2, this framework comprises the following layers [12]:

1. **The Hardware Platform Layer (HPL)** is used to classify all relevant wireless technologies.

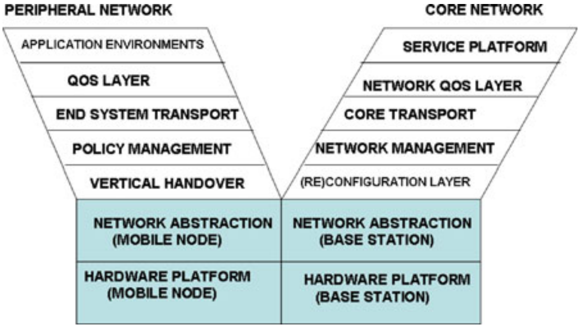


Fig. 2 The Y-Comm framework

2. **The Network Abstraction Layer (NAL)** provides a common interface to manage and control all the wireless networks. These first two layers for both frameworks are similar in functionality.
3. **The Vertical Handover Layer (VHL)** acquires the resources for handover.
4. **The Policy Management Layer (PML)** decides whether and when handover should occur.
5. **The End Transport Layer (ETL)** allows the mobile node to make end-to-end connections across the core network.
6. **The QoS Layer (QL)** monitors the QoS used by the wireless network as a whole to ensure stable operation.
7. **The Applications Environments Layer (AEL)** specifies a set of objects, functions and routines to build applications which make use of the framework.

3.2 The Core Framework

As previously mentioned, the first two layers of the Core Framework are engaged in controlling base-station operations. The third layer is called the **Reconguration Layer (REL)**. It is a control plane to manage key infrastructure using programmable networking techniques. **The Network Management Layer (NML)** is a management plane that is used to control networking operations in the core. **The Core Transport System (CTS)**, is concerned with moving data through the core network. **The Network QoS Layer (NQL)** is concerned with QoS issues within the core network. **The Service Platform Layer (SPL)** allows services to be installed on various networks at the same time.

3.3 A Generic Network Architecture for Heterogeneous Networks

Due to the fact that various networking technologies will coexist in NGNs, its network infrastructure will be owned by different operators. Additionally, new operators could install their network hardware and join the core network. However, interoperability between different operators is a key challenge in this open, heterogeneous environment. To address this issue, the study of the Y-Comm group [13] and Daidalos II [14] proposed the concept of Core End-Points (CEPs) as administrative entity to control the different wireless networks in a regional area, as shown in Fig. 3.

A detailed view of the CEP's structure along with the attached networks is shown in Fig. 4. The figure shows a hierarchical architecture, where the bottom level is represented by several access points (APs) and access routers (ARs) that communicate with the wireless interfaces in the mobile terminals. The middle level comprises a number of technology-specific domains, where each domain represents a certain network operator and technology such as 2G, 3G, and Wi-Fi. For these domains to interoperate, the CEP, which is residing at the top level acts as a central administrative domain to control the inter-domain functions and provide overall management.

In order to deal with the QoS and security tasks in this architecture, a number of operational entities have been proposed as follows:

1. **The central A3C server (CA3C):** This is the central authentication, authorization; accounting and cost (A3C) server in the CEP. The CA3C holds the service level of agreements (SLAs) along with the network level of agreements (NLAs), which describe the clients' terms for using the service and accessing networks, respectively

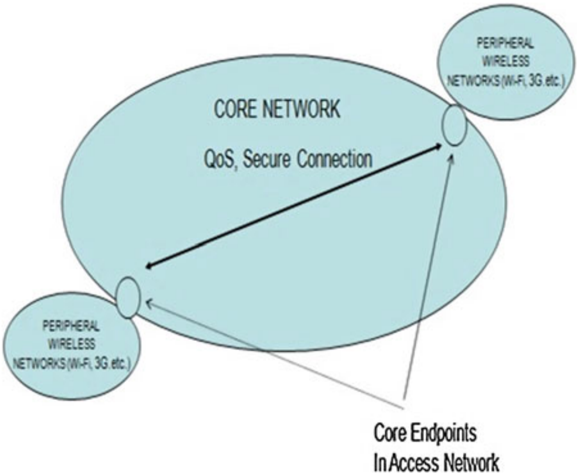

PERIPHERAL WIRELESS NETWORKS (Wi-Fi, 3G etc.)
CORE NETWORK
QoS, Secure Connection
PERIPHERAL WIRELESS NETWORKS (Wi-Fi, 3G etc.)
Core Endpoints
In Access Network

Fig. 3 The future internet

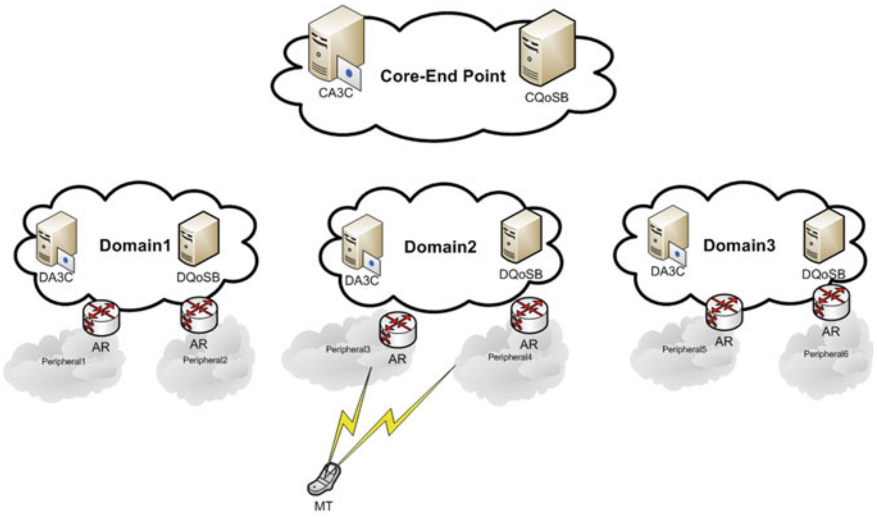


Fig. 4 The future network architecture

2. **The central QoS broker (CQoSB):** It is responsible for negotiating QoS in case of cross-CEP handover. It comprises three modules: the QoS Engine manages inter-domain connection and provides end-to-end QoS across CEPs, the A3C interface is used for the interaction with the CA3C server.
3. **The domain A3C server (DA3C):** The DA3C is responsible for handling users' service aspects.
4. **The domain QoS broker (DQoSB):** It manages the resources of the attached peripheral networks with respect user preferences and network availability, it also makes a per-flow admission control decision. The DQoSB has five modules as shown in 5, detailed description of these modules is found in [15].
5. **The access router (AR):** This is the link between the domain and the peripheral networks; it enforces the admission control decision, taken by the DQoSB. The AR comprises five modules as shown in 5, these are explained in [15].
6. **The Mobile Terminal (MT):** A multi-homed mobile device used by the subscribers to switch between different access networks to get various services. As shown in [15], the MT has four interfaces.

The the structure of these entities along with their interactions are shown in Fig. 5. Also more detailed information about these entities could be found in [15, 16].

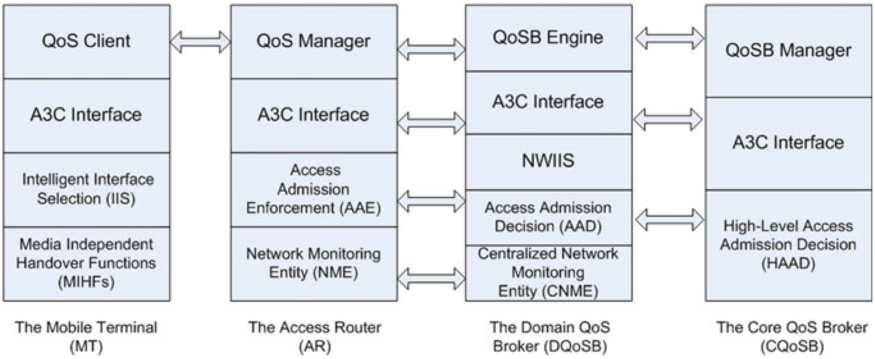

QoS Client
A3C Interface
Intelligent Interface Selection (IIS)
Media Independent Handover Functions (MIHFs)
The Mobile Terminal (MT)
QoS Manager
A3C Interface
Access Admission Enforcement (AAE)
Network Monitoring Entity (NME)
The Access Router (AR)
QoSB Engine
A3C Interface
NWIIS
Access Admission Decision (AAD)
Centralized Network Monitoring Entity (CNME)
The Domain QoS Broker (DQoSB)
QoSB Manager
A3C Interface
High-Level Access Admission Decision (HAAD)
The Core QoS Broker (CQoSB)

Fig. 5 The network entities interactions

4 Mapping Y-Comm and LTE

4.1 Mapping Y-Comm onto LTE Infrastructure

In this section we show the relationship between Y-Comm and LTE infrastructure. We believe that Y-Comm can easily be mapped onto well-established networks such as the UMTS or LTE architecture. The mobile node runs the LTE protocol stack while the required network functionality is distributed using several core entities. The eNBs interact directly with the mobile node using specified radio channels defined by the MME. Therefore, the functionality of eNBs/MMEs is mapped to the first two layers (HPL and NAL) of the Core Framework in Y-Comm. Each eNB/MMR pair is connected to the S-GW using S1 interfaces as shown in Fig. 1. The S-GW is responsible for managing mobility with other 3GPP networks such UMTS. However, the connection with packet switching networks such the Internet or supporting the mobility with non-3GPP networks such as WiMAX and Wi-Fi is achieved by the P-GWs which are connected to the S-GWs using S5/S8 interfaces. This implies that S-GWs and P-GWs are responsible for managing different types of handover as described in Sect. 2.2.1. Therefore, the functionalities of S-GWs and P-GWs correspond to the Reconfiguration Layer (REL)and the Network Management Layer (NML), respectively in the Core Framework of Y-Comm. We can now show how the functions of Y-Comm can be mapped onto the LTE infrastructure making possible the transition from LTE to Y-Comm. This is shown in Fig. 6. The Mobile Node (MN) runs the entire Peripheral Framework as shown. The Core Framework is distributed throughout the core network in a similar way to the LTE infrastructure. The Hardware Platform and Network Abstraction Layers run in the eNBs and MME. Y-Comm however, supports different wireless technologies including 3G base stations, Wi-Fi and WiMax APs, etc. The Reconfiguration Layer of Y-Comm runs in the S-GW for LTE. This layer uses programmable techniques on the Network Abstraction Layer to control

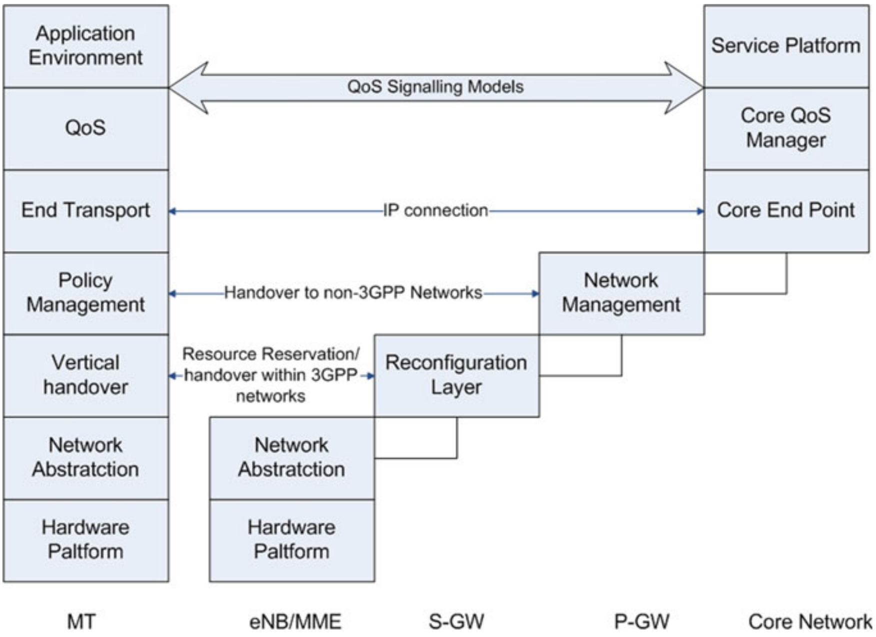


Fig. 6 Mapping the LTE to Y-Comm

the resources on individual eNBs. The Reconfiguration Layer on the S-GWs level and allocates resources to do a handover to a particular eNB.

The Network Management Layer (NML) manages different wireless networks and runs at the level of the P-GW level in current LTE infrastructure. In Y-Comm, a local NML manages all the S-GWs in a local area and knows the status of each wireless network and its topology. This information can be shared with the Policy Management Layer on the mobile node. The core endpoint is used by the mobile node to connect to the wider Internet. For a given connection, IP packets to and from the mobile node are tunnelled through the core network using core endpoints. Finally when an application on the mobile node wishes to make a connection through the core network, the QoS layer running on the mobile node interacts with the QoS manager in the core network with regard to QoS requirements for the new connection. The QoS manager will return two core endpoints which can be used for the new connection.

4.2 Mapping the LTE Infrastructure to the Generic Network Structure

Both Y-Comm and LTE have defined their own network architecture along with a number of operational network entities as explained in Sects. 2.2.1 and 3.3. This

Table 1 The Y-Comm and LTE mapping

The LTE network element	The Y-Comm network entity
The Home Subscriber Server (HSS)	The Central A3C Server (CA3C) and the Central QoS Broker (CQoSB)
The Policy Control and Charging Rules Function (PCRF)	The High-level Access Admission Decision module (HAAD) of the CQoSB
The Policy Control Enforcement Function (PCEF)	The Access Admission Decision (AAD) and the Centralized Network Monitoring Entity (CNME) modules of the DQoSB
The PDN Gateway (P-GW)	The Domain QoS Broker (DQoSB) and the Domain A3C (DA3C) server
The Serving Gateway (S-GW)	The Access Admission Enforcement (AAE) and Network Monitor Entity (NME) modules of the Access Router
The eNB and the MME	The access router

section will show how the LTE networking entities are mapped to those in the generic network architecture of the Y-Comm framework.

The Home Subscriber Server (HSS) in LTE holds the subscription information of all clients, which includes QoS, security and roaming restrictions. In Y-Comm, such information is kept by the Central A3C Server (CA3C) and the Central QoS Broker (CQoSB) which reside in the Core-End Points (CEPs) as seen in Fig. 4.

The Policy Control and Charging Rules Function (PCRF) in the core network and the Policy Control Enforcement Function (PCEF),which resides in the P-GW are responsible for policy control decision-making and controlling the flow based on the derived policy. In Y-Comm, these functions are achieved using the High-level Access Admission Decision module (HAAD)in the CQoSB, the Access Admission Decision (AAD) and the Centralized Network Monitoring Entity (CNME) modules in the DQoSB, respectively.

The PDN Gateway (P-GW) manages mobility with non-3GPP networks as well as QoS control and flow-based charging according to the rules from the PCRF. In Y-Comm, these functionalities are the responsibilities of the Domain QoS Broker (DQoSB) and the Domain A3C (DA3C) server.

The Serving Gateway (S-GW) performs administrative duties such as collecting information for charging and monitoring the volume of data sent/received from users. These duties are accomplished via the Access Admission Enforcement (AAE) and Network Monitor Entity (NME) modules of the Access Router in the generic network architecture of Y-Comm.

The functionalities of the eNBs and their controlling unit, the MME, is delivered by the Access Router (AR) in the generic network architecture. This mapping is shown in Table 1.

5 Conclusion

This work describes two current research trends to enhance communication in future networks. While one trend is keen on enhancing current communication technologies and developing new ones, the other introduces novel communication architectures within which, various wireless technologies will be supported. The paper is the first to discuss how the two trends could be mapped and complement each other in order to enhance the research efforts towards advanced technologies in future networks.

References

1. The LTE/LTE Advanced Guidea semi-annual publication on LTE/LTE Advanced, http://lteportal.com/LTE_Business_Guide
2. Aiash M, Mapp G, Lasebae A, Phan R, Augusto M, Vanni R, Moreira E (2011) Enhancing naming and location services to support multi-homed devices in heterogeneous environments. In: Proceedings of the CCSIE 2011, London, New York, pp 25–27
3. Abramowicz H, Malmgren JG, Sachs UC, Horn Prehofer C, Karl H, Niebert N, Schieder A (2004) Ambient networks: an architecture for communication networks beyond 3g, IEEE wireless communications 11
4. Okubo A, Sakakura T, Shimizu K, Adachi F, Kuroda M, Inoue M (2004) Scalable mobile ethernet and fast vertical handover. In: Proceedings of IEEE wireless communications and networking conference, Vol 2. Atlanta, USA, pp 659–664
5. IEEE802.21, Ieee 802.21/d8.0: Draft standard for local and metropolitan area networks: media independent handover services, 2007
6. Aiash M, Mapp G, Lasebae A, Loo J, Sardis FC-W, Phan R, Augusto M, Moreira E, Vanni R (2012) A survey of potential architectures for communication in heterogeneous networks. In: IEEE wireless telecommunications symposium, 2012
7. Korhonen J, Soininen J, Patil B, Savolainen T, Bajko G, Iisakkila K (2012) IPv6 in 3rd Generation Partnership Project (3GPP) Evolved Packet System (EPS). Request for Comments: 6459. http://tools.ietf.org/html/rfc6459
8. Alcatel.Lucent.: LTE Mobile Transport Evolution. Strategic White paper. http://lte.alcatel-lucent.com/locale/en_us/downloads/Alcatel-Lucent_LTE_Transport_WhitePaper.pdf
9. Tsay J-K, Mjlsnes SF (2012) Computational security analysis of the UMTS and LTE authentication and key agreement protocols. Computer Network Security, Vol 7531. LNCS, pp 65–76
10. Chandra P (2005) Bulletproof wireless security: GSM, UMTS, 802.11 and ad hoc security. Newnes, Oxford, pp 129–158
11. ITU-T, Global information infrastructure internet protocol aspects and next generation networks, y.140.1, International Telecommunication Union. ITU- T, 2004
12. Mapp G, Shaikh F, Aiash M, Porto Vanni R, Augusto M, Moreira E (2009) Exploring efficient imperative handover mechanisms for heterogeneous wireless networks. In: International symposium on emerging ubiquitous and pervasive systems (EUPS-09), Indianapolis, Ind, USA
13. The Y-Comm Research Group. http://www.mdx.ac.uk/research/areas/software/ycommresearch.aspx
14. Almeida M, Corujo D, Sargento S, Jesus V, Aguiar R (2007) An end-to-end QoS framework for 4G mobile heterogeneous environments. In: OpenNet workshop. Diegem, Belgium
15. Aiash M, Mapp G, Lasebae A (2011) A QoS framework for heterogeneous networking. In: Lecture notes in engineering and computer science: proceedings of the world congress on engineering 2011, WCE 2011, pp 6–8

16. Aiash M, Mapp G, Lasebae A, Phan R, Loo J (2012) A formally veried AKA protocol for vertical handover in heterogeneous environments using Casper/FDR. EURASIP J Wirel Commun Network 2012:57. doi:10.1186/1687-1499-2012-57,2012.OpenSpringer

A Call Admission Control Scheme for Cellular Network to Handle Sudden Influx in a Confined Area

Bhattacharya Adrija and Choudhury Sankhayan

Abstract Call Admission Control (CAC) is a typical problem in cellular network and needs more attention for up-gradation of the existing ones. The challenge is to allow more calls without disturbing the quality of service through efficient bandwidth distribution amongst cells. Several statistical approaches are used to guess the channel requirements in near future and re-distribute the bandwidths in each cell accordingly. But all these proposed method fails, if there is a sudden influx of user's concentration in a confined area and that is happening frequently in a real scenario like in an auditorium or in a football ground. The intention of a service provider is to increase the number of allowable calls that are not being permitted by the existing schemes and thus demands a new mechanism to adapt this unpredictable situation. In this paper, a policy is proposed to manage huge data traffic within a confined area network with higher user concentration. Here the users are being prioritized based on some proposed metrics and the scheme is allowing more calls according to the priority of the competing users. The proposed concept is being implemented and also compared with an existing widely accepted scheme. The offered solution is expected to perform better than the existing CAC schemes for the above mentioned situation.

Keywords Call admission control · Cellular network · Bandwidth distribution in mobile network · Bandwidth optimization

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.

B. Adrija (✉) · C. Sankhayan
Department of Computer Science and Engineering, University of Calcutta,
Kolkata, India
e-mail: adrija.bhattacharya@gmail.com

C. Sankhayan
e-mail: sankhayan@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_14,

1 Introduction

The facilities provided by the wireless telecommunication become an integral part of our day-to-day life. In mobile communication, bandwidth distribution amongst the cells is always a crucial problem. The challenge is to propose a scheme for efficient distribution of the most costly resource bandwidth among the cells without disturbing the quality of service. Classically this problem is termed as Call Admission Control (CAC) in cellular network. Inefficient distribution of bandwidth may create congestion, though there are plenty of resources available in other cells at that time. It needs a dynamic distribution scheme that can access the density of nodes in each cell and will be able to change the distribution scheme accordingly.

A new call can be admitted in a cell depending on the available bandwidth at that time instant. Either the call would be allowed or blocked due to the lack of resources. The situation for ongoing call is a little bit different [4]. A node (with running call) may move from one cell to another (handover) and due to the lack of resources the call may be dropped. Generally an ongoing call should get more priority than a new call. The performance of a call admission control scheme depends on the call blocking and call dropping probabilities. Hence the effective distribution of bandwidth may allow more calls, minimizing the number of call dropped and call blocked that leads to a good call admission control scheme.

There are several call admission control mechanisms for bandwidth allocation to users in fair way and to manage the incoming calls in an efficient manner. Several mathematical and Statistical procedures are also applied to handle the situation and to ensure comparatively better allocation [2]. But the problem is of more severe kind when there is sudden influx in number of users within a confined area. That results difficulty to manage the sudden increase in traffic load within the network. Moreover it leads to technical adversity like insufficient bandwidth and high network interference. Thus it remains a challenge to the CAC solution provider to propose a scheme that will be able to perform better than the existing schemes in this kind of unpredictable situation. The most crucial intention of the service providers are to allow as many as users and to ensure the services at least for some priority users in this typical situation. In this paper, an attempt is made to achieve the above mentioned objectives.

2 Related Work

In this section, a brief outline of the dominant existing approaches for call admission control is presented. A few of existing CAC schemes are briefly discussed here to clarify the scenario and the necessity of proposing a new CAC.

First come first serve is the most common and simple solution that can be proposed to solve the problem of call admission control. In this approach, [5] the total bandwidth is divided into segments and each segment can serve only a particular type of call. The main disadvantage here is the wastage of bandwidth due to the static boundary concept. In order to overcome this difficulty the idea of movable boundary scheme [3] is introduced where the segment widths are adjustable. But it can not provide priorities to new call and handoff call separately which is a very important objective in proposing CAC. This could be improved using the concept of guard channel where the new and handoff calls are treated differently. Using the concept of dynamic partitioning scheme [7], some channels are reserved exclusively for voice calls and data calls and the remaining channels are shared for both voice and data calls. A new voice call or handoff voice call is accommodated either in the reserved channels or in shared channels depending on the availability. The wastage of resources in this scheme is obvious as there may be a situation when the bandwidth would be available for data call but not for voice call. In order to overcome this disadvantage of dynamic partitioning scheme the Dual Threshold Bandwidth Reservation (DTBR) [6] was introduced. It has higher network utilization and provides better quality of service. In DTBR, the channels of the cell are divided into regions by thresholds. The system will drop handoff voice call if the channels are unavailable. The concept of reserve channel strategy along with DTBR is proposed to overcome this problem [8]. The combination of DTBR and reserve channel CAC scheme offer much better channel efficiency and lower call-blocking and call-dropping probabilities. Markov chain [8] is used to calculate call dropping and call blocking probabilities in R-DTBR. The quality of services in case of CAC schemes is often measured by the handoff call dropping probability and that is better in [8].

A model has been discussed in [9] to predict the amount of resource demands directly in real time, based on time series analysis. This scheme is working fine in general situations but may not be efficient to handle a sudden influx.

The major objective of a call admission control policy is to decide whether to take a particular call request or not. In order to take such a decision at a particular time, some factors are considered that are helpful to maximize network utilization and minimize rejection ratio. Incorporating the concept of priority may be helpful in this context. Assigning a priority to the calls will guarantee lower call rejection ratio for high priority user. Call requests are classified into categories based on customer type [1]. Here it is assumed that arrival time, call duration and required bandwidth are different for each customer type. Required bandwidth is predicted using statistical method. Let us consider there are M categories. A numbers of thresholds (Ti) are being fixed for each of the categories ($\mathrm{C_i}$). When a call request from Category i ($\mathrm{C_i}$) arrives at the base station and the required bandwidth of the call request is less than the remaining bandwidth of the system and as well as of $\mathrm{T_i}$, then only the call request is accepted. Further, if the required bandwidth of the call request is less than the remaining bandwidth of the system but more than T_i, then the call request is admitted to the buffer if there is no call request from the $\mathrm{C_i}$ already. Otherwise the request is rejected. The solution is not satisfactory for the particular confined area network as any call request cannot be accepted until a channel allotted for the particular category of calls is available. Thus, though there may exist a free channel but cannot be assigned to any call unless the call type is matching with that of the free channel.

3 Proposed Solution

In this section, a new CAC scheme is proposed based on statistical prediction methodologies to ensure essential and less interrupted services to the prioritized users in this typical situation. The handoff calls are not so much concerned in this case as the concentration is mainly on a confined area with a sudden increase in user's concentration. The objective is to decrease the call blocking probability and to maintain the service as much as possible for the priority users. Users are accommodated based on their classified privilege categories. Users are classified into categories based on some proposed parameters. The high user concentration in a confined area can be handled by allocating proper weight to each of the concerned parameters involved in categorizing. Weights and corresponding levels of parameters are determined by choosing an unbiased sample from the total users. Parameters considered are Duration of communication, Frequency of each communication type, Average cost per minute, Rental paid by the user. Based on the sampled data, each parameter is divided into some levels and each level is assigned with a score. Now a combined score is calculated depending on the score of the proposed parameters for deciding the category of the concerned user. The parameters are selected carefully to reflect the user profile as these are found to have direct relation with loss or profit of a service provider. The user producing higher profit for service provider gains higher priority.

3.1 Duration of Communication

Duration of communication reflects how much amount of time a user make calls. In general, it is simple that higher the duration higher the profit of service provider. Thus duration of call is the most important metric to asses the profile of the user from the view of a service provider. Data on duration of communications of a certain user for past few months is scrutinized and based on that average call duration of a communication is predicted. A few methods are well-known for predicting values. By taking only the seasonal fluctuation under consideration a method of weighted moving average is used. In this method all previous available data is taken into account according to their relevancy that is older the data lesser the contribution in calculation. For example, the Service provider is considered to have 1000 users. User records for previous three months are available. Among which a random sample of 100 is drawn to analyze and categorize users. For each of the 100 user average duration of communication is calculated by Weighted moving average method.

3.2 Frequency of Each Communication Type

In general a particular cell phone can offer limited number of services. Here, higher the data size of the communication, higher the weights in measuring user profile.

The types of requests accepted or forwarded by a service provider vary with the type of communication of the user wants as well as the type of plan the user have selected. Initially, types of requests considered are namely Voice call, SMS, Video data transferring, Internet surfing. The score of a user is calculated based on the used amenities in the chosen sample. The transaction type for all users in the sample can be found out and then the scores are given based on the most likely call-type to occur.

3.3 Average Cost Per Minute

The bill amount of the user not always reflects the actual importance of the user. Higher amount of bill does not always imply that the usage is high. It is a common case that a particular user has chosen a plan that is enough cost cutting even for higher amount of data consumption and results no or little profit for service provider. Thus it is crucial to grade a particular user based only on previous bill paid. To maintain uniformity in some means the average call cost is calculated. Thus cost per minute is a measurable amount that gives the average cost per communication. Let us declare variable Costpermin. It is measured by:

$$\mathbf{Costpermin} = (\mathbf{avg\ bill\ per\ week})/(\mathbf{avg\ call\ duration\ per\ week}).$$

3.4 Choice of Rental

The idea of rental is introduced to provide user better value for money. The actual usage pattern is not reflected independently by Average cost per minute. Another parameter considered here is the choice of rental. It is evident that a customer paying lesser rental gives more profit in general from service provider's perspective. Thus both the components Average cost per minute and choice of rental helps to reflect a user profile correctly.

3.5 Choice of Weights and Final Score Calculation

The weights given to all the four factors considered previously are very important to formulate. The weights are decided According to the importance of the factors in assessing a user profile. Duration of Communication is very important in the context of confined area network with higher user concentration as lower the call duration lower the congestion and higher the call request accepted in a stipulated time. But from service provider's perspective, it is better to have long duration call costing higher profit. Though in some way shorter call duration will make the situation better, it can't be ensured that premium users always request for short calls. Thus considering average cost per minute, the privilege users are found (giving priority to the higher call cost). These two factors alone can not reflect the actual scenario. Thus two new parameter introduced but these have lesser important. Often short duration communication may transmit bulk data or the huge rental facilitates the user to send bulk data in nominal cost per minute. Communication type and rental chosen by the user are considered as additional parameters. Duration of communication and Average cost per minute should have same weights. Frequency of each communication type has the next higher weight, whereas the rental paid by the user plays the least significant role in evaluating user's profile. So the total score of a customer will be $W = \mathbf{g}_1{}^*\mathbf{w}_1+\mathbf{g}_2{}^*\mathbf{w}_2+\mathbf{g}_3{}^*\mathbf{w}_3+\mathbf{g}_4{}^*\mathbf{w}_4$, where ith parameter has the score g_i and a pre determined weight for i^{th} parameter be w_i. Now, whenever a burst of requests are coming simultaneously to the base station, it should content the user ID and the corresponding score of that user. The Base station takes decision by taking into account the score of the user and the current network (channel) occupancy.

Let us consider that the score for jth user be w_j .There exists a predefined threshold value say T (Threshold). The value of T at a specified point of time is dependent on the current network utilization. A request will be granted if the score of that user is greater than the threshold value of the network. Otherwise the call will be blocked. The threshold may be at a very low value at earlier point of time. Then all low privilege (low score) user requests can be granted. But with the increasing of load within a network, T is increasing. As a result only the higher privilege user requests are granted. The handling of call request is depicted in Fig. 1. The data for scores can

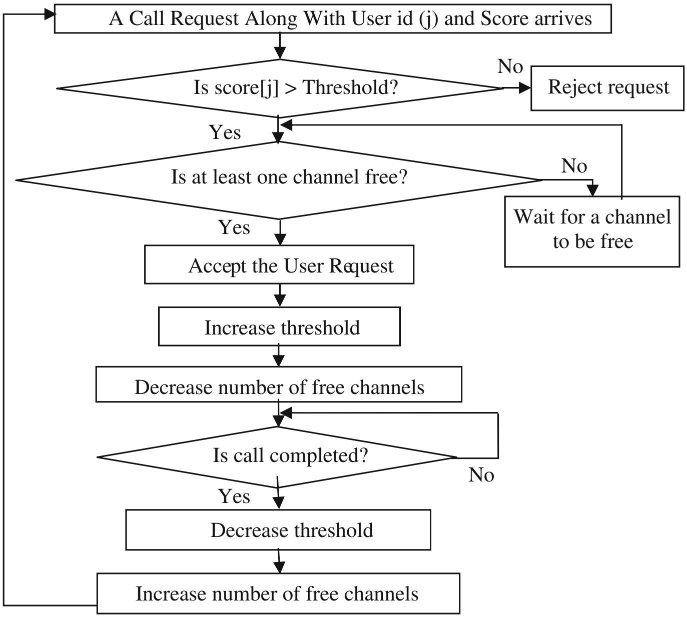

A Call Request Along With User id (j) and Score arrives
Is score[j] > Threshold?
No
Reject request
Yes
Is at least one channel free?
No
Wait for a channel to be free
Yes
Accept the User Request
Increase threshold
Decrease number of free channels
Is call completed?
No
Yes
Decrease threshold
Increase number of free channels

Fig. 1 Call handling by proposed scheme

be updated periodically (period may be a month). Each time the database updated the analysis is done on most recent data.

4 Conclusion

Primary objective of proposing a call admission control scheme was to propose a scheme to provide cellular services at high user concentrated confined area network. The confined area network will be any one of the train, bus, stadium, cinema hall and theatre etc, where a huge gathering of people is often present. It is tough for a network service provider to handle a certain influx in confined area network when a huge number of present users are simultaneously requests for communication. It is discussed so far that there is no such existing solution to the particular problem that can minimize the number of blocked users, provide satisfaction to privilege users as well as maintaining a satisfactory level of the quality of service.

In this proposed scheme, it is tried to ensure the minimum services at least to the privilege users. After doing all the data analysis work the final score of each of the 1000 users is known. Combining scores for all the parameters the final score for each of the hundred users is calculated. Further, the scheme discussed in [1] and the proposed schemes are implemented using C. There are some assumptions of the implementation such as, call requests are generated randomly, call duration of the call are also random, the channel width is fixed for both the schemes. The simulation is done for a time period of 1000 s. In the proposed scheme, privilege users get better service than that in [1]. Call request time is within 1000 s and call duration assumed to be not more than 2000 s. The number of user categories in [1] is considered to be 10. Each category has 5 dedicated channels. User Score in case of the proposed solution is assumed to be any integer value from 0 to 7. In both the cases total number of channels are assumed to be 40 and 80. The numbers of requests at a specified time interval for varying channel number are as follows: For channels = 40, total number of call requests served by scheme [1] is 175 and that of proposed scheme is 221 (Fig. 2). For channels = 80 total number of call requests served by scheme [1] is 282 and that of proposed scheme is 360 (Fig. 3).

Experimental results have shown that proposed scheme can cater more number of requests also. A comparative study is being done between the existing solution [1] and the proposed work. This CAC scheme performs better compared to the existing solutions for this specific environment as it ensures better service to privileged users and that is really an attractive scheme from the service provider point of view.

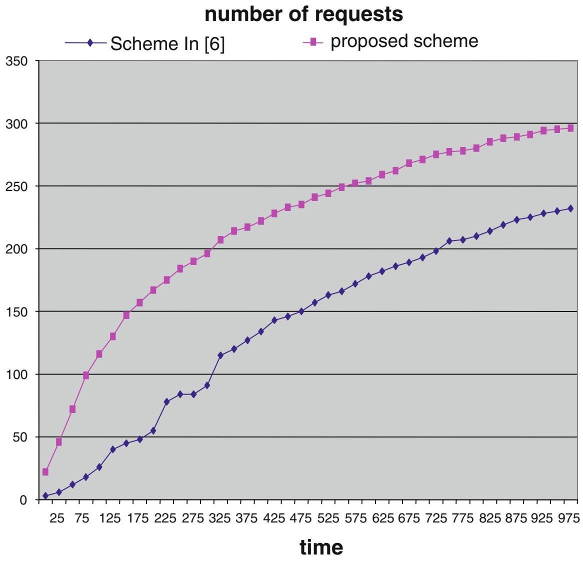


Fig. 2 Diagram showing comparisons between two schemes when total number of channels = 40

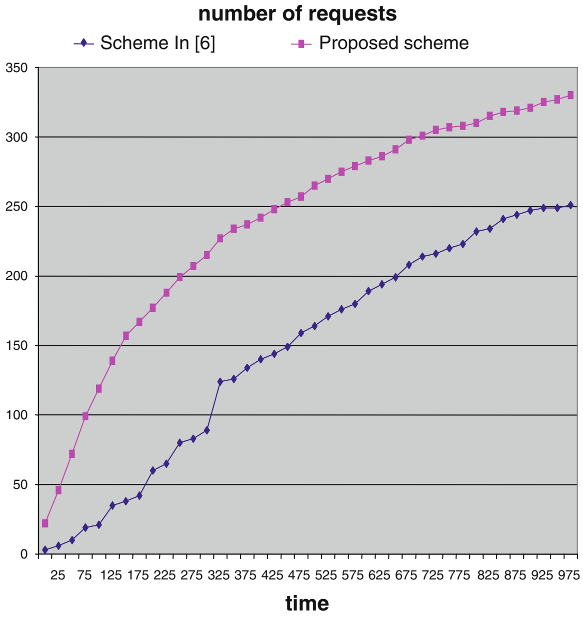


Fig. 3 Diagram showing comparisons between two schemes when total number of channels = 80

References

1. Aboleaze MA, AloulA FA (2004) Call admission protocol for wireless cellular multimedia networks. In: Vehicular Technology Conference VTC 2004-Spring
2. Ghaderi M, Boutaba R (2005) Call admission control in mobile cellular networks: a comprehensive survey. Wirel Commun Mobile Comput 6(1):69–93
3. Huang YR, Lin YB, Ho JM (2000) Performance analysis for voice/data integration on a finite mobile systems. IEEE Trans Veh Technol 49:367–378
4. Kolate VS, Sonawane BS, Bhide AS (2012) Improving QoS in 3G wireless mobile n/w using call admission control scheme. IJCSET 2(3):1016–1019
5. Lai YC, Tsai SF (2002) A fair admission control for large bandwidth multimedia applications. In: Proceedings of the GLOBECOM, Sam Antonio, TX, pp 317–322, Nov 2002
6. Li L-Z, Li B, Li B, Cao X-R (2007) Performance analysis of bandwidth allocations for multisenrice mobile wireless cellular networks. In: Tavel P (ed.) Proceedings of the IEEE WCNC, pp 1072–1077, March 2003, (2007)
7. Li B, Li L, Li B, Sivalingam KM, Cao X-R (2004) Call admission control for voice/data integrated cellular networks: performance analysis and comparative study. IEEE J Sel Areas Commun 22:706–718
8. Ng HY, KO KT, Tsang KF (2005) 3G mobile network call admission control scheme using Markov chain. In: Proceedings of the ninth international symposium on consumer electronics (ISCE 2005), pp 276–80, ISBN: 0-7803-8920-4
9. Zhang T et al (2001) Local predictive resource reservation for handoff in multimedia wireless IP networks. IEEE J Sel Areas Commun 19:1932–1941

Distributed Joint Optimal Network Scheduling and Controller Design for Wireless Networks

Hao Xu and S. Jagannathan

Abstract In this paper, a novel distributed joint optimal network scheduling and controller design for the wireless network control system (WNCS) application is introduced via a cross layer design approach. First a stochastic optimal controller design that minimizes an infinite horizon optimal regulation of uncertain linear system with wireless imperfections due to wireless network protocol is proposed. Subsequently, a novel optimal cross-layer distributed scheduling scheme is presented as part of wireless network protocol design. Compared with traditional scheduling schemes, the proposed cross-layer distributed scheduling scheme can not only optimizes the utility function of wireless network but also satisfies controller demands on packet loss probability and delay bounds in order to main the overall system stable. Simulation results are included to illustrate the effectiveness of the proposed cross layer co-design.

Keywords Distributed scheduling · Cross layer · Wireless networked control system · Utility-optimal

1 Introduction

In contrast with traditional dedicated control systems, wireless network is now being utilized in control systems making the systems distributed. These novel distributed systems are known as wireless networked control system (WNCS) [1–3]. Practical examples of such systems include smart power grid, network enabled manufacturing,

H. Xu (✉) · S. Jagannathan
Department of Electrical and Computer Engineering,
Missouri University of Science and Technology, Rolla, MO, USA
e-mail: hx6h7@mst.edu

S. Jagannathan
e-mail: sarangap@mst.edu

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_15,

water distribution, traffic, and so on. In WNCS, wireless communication packets carry sensed data and control commands from different physical systems (or plant) and remote controllers. Although WNCS can reduce system wiring, ease of system diagnosis and maintenance, and increase agility, the uncertainty caused by shared wireless network and its protocols and practical natural of control systems bring many challenging issues for both wireless network protocol and controller designs.

The issues for control design [2] include wireless network latency and packet loss [3] which dependent on wireless communication channel quality and the network protocol design respectively. In general, wireless network latency and packet losses can destabilize the real-time control system and in many cases can result in safety concerns. However, the recent literature [1–3] only focused on stability or optimal controller design by assuming wireless network latency and packet losses to be a constant or random ignoring the real behavior of the wireless network component. By contrast, the current wireless network protocol designs [4, 5] ignore the effects of the real-time nature of the control system or the application making them unsuitable for real-time control applications. Thus, a truly cross-layer WNCS co-design which not only optimizes the performance of the wireless network but also controlled system is necessary. Towards this end, the distributed scheduling scheme is critical in wireless network protocol design [4]. Compared with traditional centralized scheduling [5], the main advantage with distributed scheduling is that it does not need a central processor to deliver the schedules after collecting information from all the users. In IEEE 802.11 standard [4], carrier sense multiple access (CSMA) protocol is introduced to schedule wireless users in a distributed manner where a wireless node wishing to transmit does so only if it does not hear an on-going transmission. Meanwhile, fairness is a non-negligible factor in distributed scheduling design. In [6], authors proposed distributed fairness scheduling in packet switched network and wireless LAN network respectively. Different users wishing to share wireless channel can be allocated bandwidth in proportion to their weights, which ensured the fairness among different users [6]. However, since random access scheme is used in most CSMA-based distributed scheduling [6, 7] and these schemes [6, 8, 9] focus on improving the performance of link layer alone which in turn increases wireless network latency, these protocols are both not optimal and unsuitable for WNCS since they can cause degradation in the performance of WNCS.

Thus, this paper proposes a novel cross-layer approach which considers controller information from application layer and wireless network performance from MAC layer to derive a stochastic optimal controller and distributed scheduling schemes for WNCS. Proposed stochastic control design can generate the optimal control inputs through value function estimator by relaxing system dynamics, wireless network latency and packet losses. On the other hand, proposed cross-layer distributed scheduling protocol optimizes not only the wireless network performance but also the controlled plant (or system) performance by maximizing utility function generated from both the wireless network and the plant.

2 Background of Wireless Networked Control System

The basic structure of WNCS with multiple users is shown in Fig. 1 where users include controller-plant pair (or system) sharing a common wireless network. Each WNCS pair includes five main components: (1) Real-time physical system or plant to be controlled; (2) A sensor measures the system outputs from the plant; (3) A controller that generates commands in order to maintain a desired plant performance; (4) A wireless network to facilitate communication among plants with their controllers and (5) the actuators which change plant states based on commands received from the controller. Figure 2 illustrates the structure of a WNCS pair. It is important to note that $\tau_{sc}(t)$ represents wireless network latency between the sensor and the controller, $\tau_{ca}(t)$ is wireless latency between the controller and actuator and $\gamma(t)$is the packet loss indicator.

Since control system and wireless network can affect each other, a co-design approach is introduced in the next section to jointly optimize the control system and wireless network based on the information from both controlled plant information from application layer and shared wireless network information in MAC layer via the cross-layer approach. The proposed WNCS co-design includes two tasks: (1) stochastic optimal control design in application layer; and (2) a novel optimal cross-layer distributed scheduling algorithm design for MAC layer.

Fig. 1 Multiple WNCS pairs

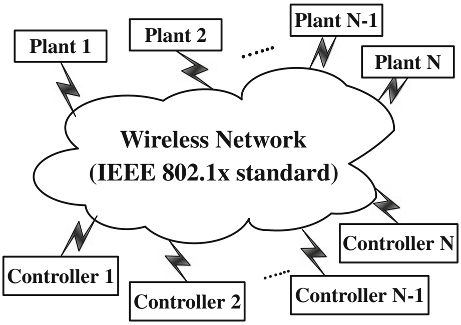
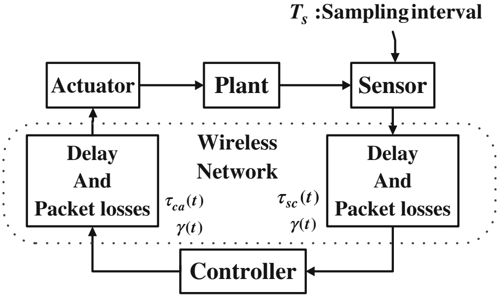


Fig. 2 Structure of one WNCS pair



3 Wireless Networked Control System (WNCS) Co-Design

3.1 Overview

In our algorithm, the control design and scheduling protocols are implemented into all WNCS pairs which are sharing the wireless network. Each WNCS pair tunes its stochastic optimal controller under wireless imperfections (e.g. wireless network latency and packet losses) caused by current distributed scheduling design, estimates its value function [10] based on tuned control design, and transmits that information (i.e. value function) to the link layer. The link layer tunes user's distributed scheduling scheme based on throughput from the link-layer and the value function value received from the application layer (i.e. controlled plant). The cross-layer WNCS co-design framework is shown in Fig. 3.

3.2 Plant Model

Suppose each WNCS pair is described as a linear time-invariant continuous-time system $\dot{x}^l(t) = A^l x(t) + B^l u^l(t) \forall l = 1, 2, ..., N$ with system dynamics A^l, B^l for lth WNCS pair, and sampling interval is T_s. For stochastic optimal control design, the wireless network latency for every WNCS pair has to be bounded as $\tau^l \leq \overline{d} T_s, \forall l = 1, \ldots, N$ which needs to be ensured by the proposed cross-layer

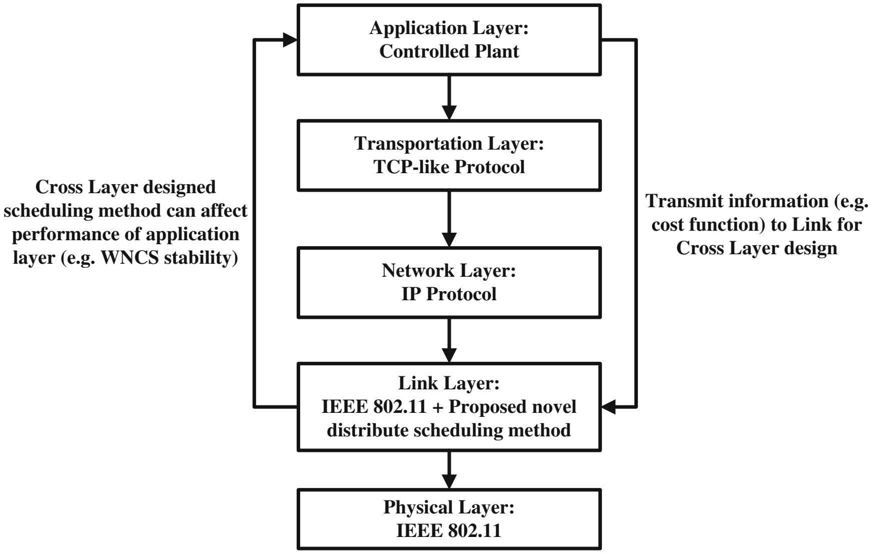

Application Layer:
Controlled Plant
Transportation Layer:
TCP-like Protocol
Network Layer:
IP Protocol
Link Layer:
IEEE 802.11 + Proposed novel distribute scheduling method
Physical Layer:
IEEE 802.11
Cross Layer designed scheduling method can affect performance of application layer (e.g. WNCS stability)
Transmit information (e.g. cost function) to Link for Cross Layer design

Fig. 3 Network structure for cross-layer design

distributed scheduling protocol. Considering wireless network latency and packet losses, the lth WNCS pair dynamics can be represented as [11]:

$$\begin{aligned} x_{k+1}^{l} &= A_s^l x_k^l + B_k^{l1} u_{k-1}^{la} + B_k^{l2} u_{k-2}^{la} + \cdots B_k^{l\overline{d}} u_{k-\overline{d}}^{la} + B_k^{i0} u_k^{la} \\ u_{k-i}^{la} &= \gamma_{k-i}^l u_{k-i}^l \qquad \forall i = 0, 1, 2, ..., \overline{d}, \forall k = 0, 1, 2.... \end{aligned} \tag{1}$$

where u_k^{la} is the actual control input received by lth WNCS actuator at time kT_s, u_k^l is the control input computed by lth WNCS controller at time kT_s, and stochastic variables γ_k^l models the packet losses for lth WNCS at time kT_swhich follows Bernoulli distribution with $P(\gamma_k^l = 1) = \overline{\gamma}^l$. A_s^l, B_k^{l1}, ..., $B_k^{l\overline{d}}$ are the augment system model dynamics caused by wireless network latency and packet losses for lth WNCS at time kT_s (Note: the definition of these dynamics is given in our previous paper [11], due to page limitation details are neglected here). By defining the augment state $z_k^l = [(x_k^l)^T (u_{k-1}^l)^T (u_{k-2}^l)^T \cdots (u_{k-\overline{d}}^l)^T]^T$ of lth WNCS pair at time kT_s, the plant dynamics (1) can be rewritten as

$$z_{k+1}^l = A_{zk}^l z_k^l + B_{zk}^l u_k^l \tag{2}$$

where the time-varying augmented system matrices are given

$$A_{zk}^l = \begin{bmatrix} A_s^l & \gamma_{k-1}^l B_k^{l1} & \cdots & \gamma_{k-i}^l B_k^{lk} & \cdots & \gamma_{k-\overline{d}}^l B_k^{l\overline{d}} \\ 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & I_m & \cdots & \cdots & 0 & 0 \\ \vdots & 0 & I_m & \cdots & \cdots & 0 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & \cdots & I_m & 0 \end{bmatrix}, B_{zk}^l = \begin{bmatrix} \gamma_k^l B_k^{l0} \\ I_m \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

It is important to note that there are two main challenges in (2) for the co-design. In practical WNCS, since wireless imperfections are not known before hand, system representation (2) is uncertain, and optimal control for WNCS has to be designed without knowing the system dynamics which is the first challenge for the control part of the co-design. Second, stochastic optimal control is designed based on the constraints of wireless imperfections. However, these wireless imperfections depend upon wireless network protocol and scheduling scheme. Therefore designing an optimal distributed scheduling protocol which not only optimizes wireless network performance but also satisfies wireless network imperfection constraints from different WNCS pairs is another challenge for the co-design.

Based on these challenges, stochastic optimal controller and cross-layer distributed scheduling schemes are proposed.

3.3 Stochastic Optimal Control

In this part, a novel stochastic optimal control is proposed for uncertain plant dynamics and wireless imperfections. Without loss of generality, lth WNCS pair is chosen for convenience for the optimal control development.

Based on optimal control theory [11], lth WNCS stochastic value function can be defined as

$$V_k^l = \underset{\tau,\gamma}{E}\{(z_k^l)^T P_k^l z_k^l\} \tag{3}$$

where $P_k^l \geq 0$ is the solution of Stochastic Riccati Equation (SRE) for lth WNCS pair and $\underset{\tau,\gamma}{E}\{\bullet\}$ is expect operator (in this case the mean value) of $\{(z_k^l)^T P_k^l z_k^l\}$. Stochastic optimal control for lth WNCS pair at time kT_s can be solved by minimizing value function, i.e. $(u_k^l)^* = \arg\min V_k^l \forall k = 1, 2, ...$ Similar to [11], the value function (3) can be expressed as

$$V_k^l = \underset{\tau,\gamma}{E}\{[(z_k^l)^T, u_k^{lT}]H_k^l[(z_k^l)^T, u_k^{lT}]^T\} = (\overline{h}_k^l)^T \overline{\chi}_k^l \tag{4}$$

where

$$\overline{H}_k^l = \underset{\tau,\gamma}{E}(H_k^l) = \begin{bmatrix} \overline{H}_k^{lzz} & \overline{H}_k^{lzu} \\ \overline{H}_k^{luz} & \overline{H}_k^{luu} \end{bmatrix}$$
$$= \begin{bmatrix} S_z^l + \underset{\tau,\gamma}{E}[(A_{zk}^l)^T P_{k+1}^l A_{zk}^l] & \underset{\tau,\gamma}{E}[(A_{zk}^l)^T P_{k+1}^l B_{zk}^l] \\ \underset{\tau,\gamma}{E}[(B_{zk}^l)^T P_{k+1}^l A_{zk}^l] & R_z^l + \underset{\tau,\gamma}{E}[(B_{zk}^l)^T P_{k+1}^l B_{zk}^l] \end{bmatrix}$$

$\overline{h}_k^l = vec(\overline{H}_k^l)$, $\chi_k^l = [(z_k^l)^T u^T (z_k^l)]^T$ and $\overline{\chi}_k^l$ is the Kronecker product quadratic polynomial basis vector of lth WNCS pair and $\overline{h}_k^l = vec(\overline{H}_k^l)$ with the vector function acting on a square matrices thus yielding a column vector (Note: the $vec(\bullet)$ function is constructed by stacking the columns of the matrix into one column vector with the off-diagonal elements which can be combined as $H_{mn}^l + H_{nm}^l$ [11]). According to [10], the optimal control of lth WNCS pair can be expressed by using H^lmatrix as

$$(u_k^l)^* = -[R_z^l + \underset{\tau,\gamma}{E}(B_{zk}^{lT} P_{k+1}^l B_{zk}^l)]^{-1} \underset{\tau,\gamma}{E}(B_{zk}^{lT} P_{k+1}^l A_{zk}^l) z_k^l = -(\overline{H}_k^{luu})^{-1} \overline{H}_k^{luz} z_k^l. \tag{5}$$

Therefore, if H^l matrix is obtained for lth WNCS pair, then stochastic optimal control is solved. However, since system dynamics is unknown, H^l matrix cannot be solved directly. Similar to [11], we adaptively estimate the value function H^l matrix and obtain the optimal control. Value function estimation error e_{hk}^l of lth WNCS pair at time kT_s can be defined and expressed as $\hat{V}_k^l - \hat{V}_{k-1}^l + z_k^{lT} S_z^l z_k^l + u_k^{lT} R_z^l u_k^l = e_{hk}^l$, where $\hat{V}_l$ is the estimated stochastic value function of lth WNCS pair at time kT_s, and S_z^l, R_z^l are positive definite matrix and positive semi-definite matrix of lth WNCS pair respectively. Then, update law for parameter vector for the value function can be given as

$$\hat{\bar{h}}_{k+1}^l = \hat{\bar{h}}_k^l + \alpha_h^l \frac{\Delta\overline{\chi}_k^l (e_{hk}^l - z_k^{lT} S_z^l z_k^l - u_k^{lT} R_z^l u_k^l)^T}{\Delta\overline{\chi}_k^{lT} \Delta\overline{\chi}_k^l + 1} \tag{6}$$

where $\overline{\chi}_k^l$ is regression function of lth WNCS pair, $\Delta\overline{\chi}_k^l$ is defined as $\Delta\overline{\chi}_k^l = \overline{\chi}_k^l - \overline{\chi}_{k-1}^l$ and α_h^l is the learning rate of value function estimator for lth WNCS pair respectively.
Based on estimated H^l matrix, the stochastic optimal control for lth WNCS pair can be expressed as

$$\hat{u}_k^l = -(\hat{\overline{H}}_k^{luu})^{-1} \hat{\overline{H}}_k^{luz} z_k^l. \tag{7}$$

Algorithms 1 and 2 represent the proposed stochastic optimal control design while Theorem 1 shows that the value function estimation errors are asymptotically stable. Further, the estimated control inputs will also converge to the optimal control signal asymptotically.

Algorithm 1 Stochastic Optimal Control for *lth* WNCS pair

```
1: Initialize: $\hat{\bar{h}}_0^l = \mathbf{0}$ and implementing admissible control $u_0^l$
2: while { $kT_s \le t < (k+1)T_s$ } do
3:     Calculate the value function estimation error $e_{hk}^l$ .
4:     Update the parameters of the value function estimator
5:         $\hat{\bar{h}}_{k+1}^l = \hat{\bar{h}}_k^l + \alpha_h^l \frac{\Delta\overline{\chi}_k^l (e_{hk}^l - z_k^{lT} S_z^l z_k^l - u_k^{lT} R_z^l u_k^l)^T}{\Delta\overline{\chi}_k^{lT} \Delta\overline{\chi}_k^l + 1}$ .
6:     Update control input based on estimated $H^l$ matrix.
7:         $\hat{u}_k^l = -(\hat{\overline{H}}_k^{luu})^{-1} \hat{\overline{H}}_k^{luz} z_k^l$ .
8: end while
9: Go to next time interval $[(k+1)T_s, (k+2)T_s)$ (i.e. $k = k+1$ ), and go back to
   line 2.
```

Algorithm 2 Plant of lth WNCS pair

```
1: Initialize: $lth$ WNCS pair states $z_k^l$
2: while { $kT_s \le t < (k+1)T_s$ } do
3:     Receive and implement control inputs from controller;
4:     if multiple control inputs have been received at the plant
       on the same time then
5:     Apply the recent control input to the plant and discard other
       control inputs [7].
6:     else if old control input arrived after new control input being received
       at the plant then
7:         Discard old control inputs and apply newer control inputs
           to the plant.
8:         else the control input is received by plant at different time and
           keep in order then
9:             Apply the control inputs $\{u_{k-1}^l, u_{k-2}^l, ..., u_{k-\bar{d}}^l\}$ received
               during this time interval to the plant sequentially.
10:        end if
11:    end if
12: end while
15: Go to next time period $[(k+1)T_s, (k+2)T_s)$ (i.e. $k = k+1$), and go back
    to while loop (line 2).
```

Theorem 1 Given the initial state z_0^l for lth WNCS pair, estimated value function and value function vector $\hat{\bar{h}}_k^l$ of lth WNCS pair be bounded in the set **S**, let u_{0k}^l be any initial admissible control policy for lth WNCS pair at the time kT_s (1) with wireless imperfections satisfying latency constraints (i.e. $\tau < \bar{d}T_s$) caused by distributed scheduling protocol. Let value function parameters be tuned and the estimated control policy be provided by (6) and (7) respectively. Then, there exists positive constant α_h^l such that the system state z_k^l and stochastic value function parameter estimation errors $\tilde{\bar{h}}_k^l$ are all asymptotically stable. In other words, as $k \to \infty$, $z_k^l \to 0$, $\tilde{\bar{h}}_k^l \to 0$, $\hat{V}_k^l \to V_k^l$ and $\hat{u}_k^l \to (u_k^l)^* \forall l$.

3.4 *Novel Optimal Cross-Layer Distributed Scheduling Scheme*

In this section, we focus on novel utility-optimal distributed scheduling design which is mainly at the link layer. Therefore, without loss of generality, traditional wireless ad-hoc network protocol [12] is applied to the other layers. For optimizing the performance of WNCS and satisfying the constraints from proposed stochastic optimal control design in the application layer, a novel optimal cross-layer distributed scheduling algorithm is proposed here by using controlled plane information from application layer.

First, the utility function for lth WNCS pair is defined as

$$Utility_k^l = 2^{(R_l + \Delta V_k^l)}. \tag{8}$$

Since performance of controlled plant and wireless network are considered in wireless network protocol design, utility function includes two parts: (1) a value function from lth WNCS pair's controlled plant at time kT_s is $\Delta V_k^l = (z_k^{lT} S_z^l z_k^l + u_k^{lT} R_z^l u_k^l) - (z_k^{lT} S_z^l z_k^l + u_k^{lT} R_z^l u_k^l)$; (2) Throughput of lth WNCS pair, R_l, which can be represented as (9) by using Shannon theory as

$$R_l = B_{wncs} \log_2 (1 + \frac{P_l d_l^{-2}}{n_0^l B_{wncs}}) \tag{9}$$

where B_{wncs} is the bandwidth of the entire wireless network, P_l is the transmitting power of lth WNCS pair, d_l is the distance between plant and controller of lth WNCS pair and n_0^l is the constant noise dense of lth WNCS pair.

Next, the optimal distributed scheduling problem can be formulated as maximizing the following utility function

$$\begin{aligned} &\text{maxmize} \sum_{l=1}^{N} Utility_k^l = \text{maxmize} \sum_{l=1}^{N} 2^{(R_l + \Delta V_k^l)} \\ &\text{subject to}: \tau_k^l \leq \overline{d} T_s \qquad \forall l = 1, 2, ..., N; \forall k = 0, 1, 2, ... \end{aligned} \tag{10}$$

where τ_k^l is wireless network latency of lth WNCS pair at kT_s .

It is important to note that wireless network latency constraints in (10) represent the proposed optimal control design constraints. Cross-layer distributed scheduling is not only maximizes the sum of all WNCS pairs but also satisfies the wireless network latency requirement for every plant-controller pair which in turn ensures that the proposed control design can optimize the controlled plant properly.

The main idea of proposed distributed scheduling scheme is to separate transmission time of different WNCS pairs by using backoff interval (BI) [12] based on utility function in a distributed manner. The proposed distributed scheduling framework is shown in Fig. 4. For solving optimal scheduling problem (10) by different WNCS pairs, the BI is designed as

$$BI_k^l = \xi^* \frac{\sum\limits_{j=1}^{N} 2^{(R_j + \Delta V_k^j)}}{\beta_k^l \sum\limits_{j=1}^{N} 2^{(R_j + \Delta V_k^j)} + 2^{(R_l + \Delta V_k^l)}} \forall l = 1, 2, ..., N \tag{11}$$

where ξ is scaling factor and β_k^l is the balancing parameter of lth WNCS pair at time kT_s which is equal to the index of first unsent packet stored in transmission buffer of lth WNCS pair. It is important to note balancing parameter is used to satisfy latency constraints in (10), which is illustrated in Theorem 2.

Algorithm 3 Novel utility-optimal cross-layer distributed scheduling scheme

```
1: **Initialize:** The balancing parameters are initialized as $\beta_0^l = 0, \forall l = 1,2,...,N$ , and
   each WNCS pair broadcasts its utility function value $2^{(R_l+\Delta V_0^l)}$ and receives utility
   function values of other pairs to calculate the network utility function value (i.e.
   $\sum_{l=1}^{N} 2^{(R_l+\Delta V_0^l)}$).
2: **While** { $kT_s \le t < (k+1)T_s$ } **do**
3:     **Calculate** backoff interval (BI) by different WNCS pair
       $BI_k^l = \xi * \dfrac{\sum_{j=1}^{N} 2^{(R_j+\Delta V_k^j)}}{\beta_k^l \sum_{j=1}^{N} 2^{(R_j+\Delta V_k^j)} + 2^{(R_l+\Delta V_k^l)}} \forall l = 1,2,...,N$ .
4:     **Contend** wireless resource.
5:     **If** $lth$ WNCS pair has the smallest BI **then**
6:         **Schedule** $lth$ WNCS pair and transmit $lth$ WNCS pair's data through
           wireless network.
7:         **If** transmission is over, **then**
8:             **Update** the scheduled WNCS pair's balancing parameter $\beta_k^l$ .
9:             **Broadcast** the message to notify all the users that wireless channel
               is free.
10:        **end if**
11:    **else**
12:        **Update** entire wireless network utility $\sum_{l=1}^{N} 2^{R_l+\Delta V_k^l}$ and WNCS pairs'
           balancing parameters $\beta_k^i, \forall i,k$ .
13:        **Wait** for wireless channel to be free.
14:    **end if**
15:    **Update** time stamp: $t = t + BI_k^l + T_k^l$ ( $BI_k^l$ is the backoff interval of scheduled
       WNCS pair. $T_k^l$ is the transmission time of scheduled WNCS pair.)
16: **end while**
17: **Update and broadcast** utility function $2^{R_l+\Delta V_k^l}$ from all WNCS pairs.
18: **Go to** next time period $[(k+1)T_s,(k+2)T_s)$ (i.e. $k = k+1$ ), and go back
    to line 2.
```

Remark 1 Since every WNCS pair decides its schedule by using local information, proposed novel optimal cross-layer scheduling scheme is distributed. In this paper,

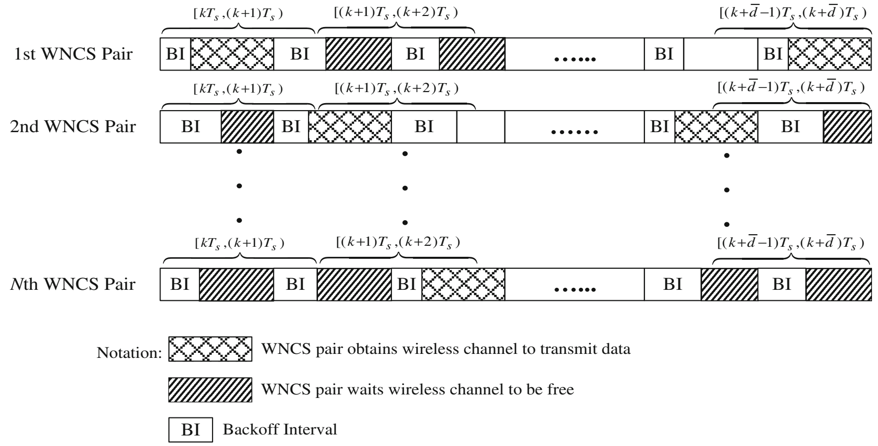

1st WNCS Pair
2nd WNCS Pair
Nth WNCS Pair
$[kT_s,(k+1)T_s)$
$[(k+1)T_s,(k+2)T_s)$
$[(k+\bar{d}-1)T_s,(k+\bar{d})T_s)$
BI
......
Notation:
WNCS pair obtains wireless channel to transmit data
WNCS pair waits wireless channel to be free
BI Backoff Interval

Fig. 4 The proposed cross-layer distributed scheduling framework

we assume that every WNCS pair broadcasts its utility function periodically in order to calculate entire wireless network utility.

Remark 2 Compared with other distributed scheduling schemes [6–8], proposed algorithm designs the backoff interval intelligently by optimizing utility function instead of selecting it randomly [6–8].

Theorem 2 The proposed distributed scheduling protocol based on cross-layer design delivers the desired performance in terms of satisfying the delay constraints $\tau_k^l \leq \bar{d}T_s \forall l = 1, 2, ..., N; \forall k = 0, 1, 2, ...$ (i.e. during$[kT_s, (k + \bar{d})T_s$), every WNCS pair should be scheduled at least once).

Proof Omitted due to page limitation.

Theorem 3 When priorities of different WNCS pairs are equal, proposed scheduling protocol can render best performance schedules for each WNCS pair.

Proof Based on the definition of BI design (11), the priority term is $\beta_k^i \sum_{j=1}^{N} 2^{(R_j+\Delta V_k^j)}$. If it is same for any WNCS pairs, it indicates

$$\beta_k^i \sum_{j=1}^{N} 2^{(R_j+\Delta V_k^j)} = \beta_k^l \sum_{j=1}^{N} 2^{(R_j+\Delta V_k^j)} \quad \forall i, l \in [1, N] \text{ and } \neq l \tag{12}$$

Therefore, for $\forall i, l \in [1, N]$ and $i \neq l$, the BI of different WNCS pair should satisfy

$$\frac{BI_k^i}{BI_k^l} = \frac{\beta_k^l \sum_{j=1}^{N} 2^{(R_j+\Delta V_k^j)} + 2^{(R_l+\Delta V_k^l)}}{\beta_k^i \sum_{j=1}^{N} 2^{(R_j+\Delta V_k^j)} + 2^{(R_i+\Delta V_k^i)}} > 1 \tag{13}$$

If and only if

$$2^{(R_l+\Delta V_k^l)} > 2^{(R_i+\Delta V_k^i)} \tag{14}$$

Next, proof for proposed cross-layer distributed scheduling algorithm is given. First of all, utility function of whole WNCS is defined as

$$Utility^{tot} = \sum_{j=1}^{N} 2^{(R_j+\Delta V_k^j)} \tag{15}$$

Assume set $\mathbf{S}_1$ is the unscheduled WNCS pairs set, and $\mathbf{S}_2$ is scheduled WNCS pairs set. $\forall l \in \mathbf{S}_1, \forall n \in \mathbf{S}_2$ such that $BI_k^l > BI_k^n$ and $2^{(R_l+\Delta V_k^l)} < 2^{(R_n+\Delta V_k^n)}$. Therefore, $U_{S2} = \sum_{j=\mathbf{S}_2} 2^{(R_j+\Delta V_k^j)}$ and U_{S2}^1 can be derived as

$$\begin{aligned} U_{S2}^1 &= \sum_{j=\mathbf{S}_2, j\neq n} 2^{(R_j+\Delta V_k^j)} + 2^{(R_l+\Delta V_k^l)} \\ &= \sum_{j=\mathbf{S_2}} 2^{(R_j+\Delta V_k^j)} + \left[2^{(R_l+\Delta V_k^l)} - (2^{(R_n+\Delta V_k^n)})\right] \sum_{j=\mathbf{S_2}} 2^{(R_j+\Delta V_k^j)} = U_{S2} \end{aligned} \tag{16}$$

Thus, when all WNCS pairs' priorities are same, the utility function of WNCS pairs based on proposed scheduling algorithm achieves maximum which illustrates the optimality.

Remark 3 Fairness is an important factor to evaluate the performance of scheduling algorithm. For proposed cross-layer distributed scheduling, a fairness index (FI) [9] is defined as $FI = \left(\sum_{i=1}^{N} \frac{R_i}{2^{(R_i+\Delta V^l)}}\right)^2 \Big/ \left[N * \sum_{i=1}^{N} \left(\frac{R_i}{2^{(R_i+\Delta V^l)}}\right)^2\right]$ to measure the fairness among different WNCS pairs.

4 Numerical Simulations

To evaluate the cross-layer co-design, the wireless network includes 10 pairs of physical plant and remote controllers which are located within 150 m*150 m square area randomly. Since batch reactor is considered as a benchmark example for WNCS [11], all 10 pairs use it. The continuous-time model is

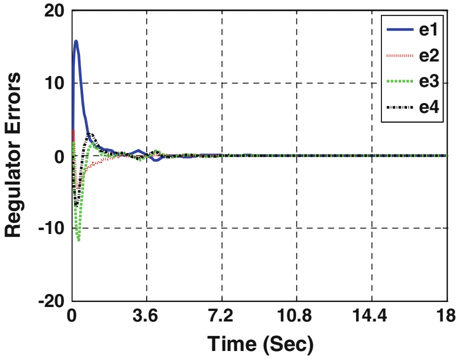


Fig. 5 State regulation errors

$$\dot{x} = \begin{bmatrix} 1.38 & -0.2077 & 6.715 & -5.676 \\ -0.5814 & -4.29 & 0 & 0.675 \\ 1.067 & 4.273 & -6.654 & 5.893 \\ 0.048 & 4.273 & 1.343 & -2.104 \end{bmatrix} x + \begin{bmatrix} 0 & 0 \\ 5.679 & 0 \\ 1.136 & -3.146 \\ 1.136 & 0 \end{bmatrix} u \qquad (17)$$

First, the performance of proposed stochastic optimal control algorithm is shown in Fig. 4. Due to page limitation, and without loss of generality, an average value of 10 different state regulation errors is shown in Fig. 5. The results indicates that the stochastic optimal control under wireless network latency with unknown dynamics and can make the state regulation errors converge to zero quickly while ensuring all WNCS stable. Note that there are some overshoots observed at the beginning because the optimal control tuning needs time.

Second, the performance of the proposed cross-layer distributed scheduling is evaluated. For comparison embedded round robin (ERR) [12] and Greedy scheduling [5] have been added. In Fig. 6, wireless network latency of each WNCS pair is shown. At the beginning, based on the different value of utility function defined in (8), one WNCS pair contends the wireless resource to communicate. Meantime,

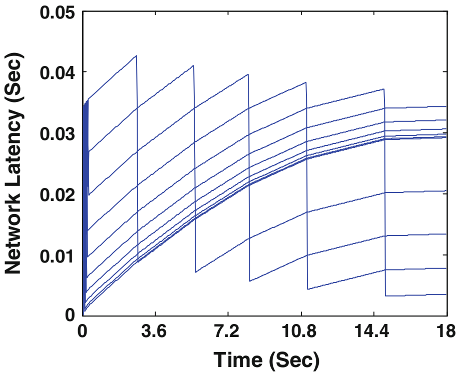


Fig. 6 Wireless network latency

since unscheduled WNCS pairs have to wait, the wireless network latency of the other WNCS pairs have been increased. However, with wireless network latencies increasing, their BI values have to be decreased and scheduled WNCS pair's BI need to be increased based on proposed cross-layer distributed algorithm (11). Therefore, when the BI of unscheduled WNCS pair is smaller than scheduled BI, it can access the wireless resource to transmit. It is important to note that wireless network latency of all 10 WNCS pairs have never been increased beyond $\overline{d}T_s$ (Note: $\overline{d} = 2, T_s = 0.034\,\text{s}$) which indicates wireless network latency constraints of all 10 WNCS pairs have been satisfied.

Third, utility function of WNCS with three different scheduling schemes is compared. As shown in Fig. 7, proposed cross-layer scheduling maintains a high value while utilities of WNCS with ERR and Greedy scheduling are much less than proposed scheduling. It indicates proposed scheduling scheme can improve the performance of WNCS better than ERR and Greedy scheduling. It is important to note since Greedy scheduling only optimize the link layer performance and utility function of WNCS is defined from both link layer and application layer, it cannot optimize the WNCS performance.

Eventually, fairness of different scheduling algorithms with different number of WNCS pairs has been compared. As shown in Fig. 8, fairness indices of proposed cross-layer distributed scheduling and ERR scheduling schemes are close and equal to one, whereas that of Greedy scheduling is much less than one thus indicating fair allocation of wireless resource for the proposed one. According to above results, the proposed cross-layer WNCS co-design optimizes the performance of both wireless network and plant.

Fig. 7 Utility comparison

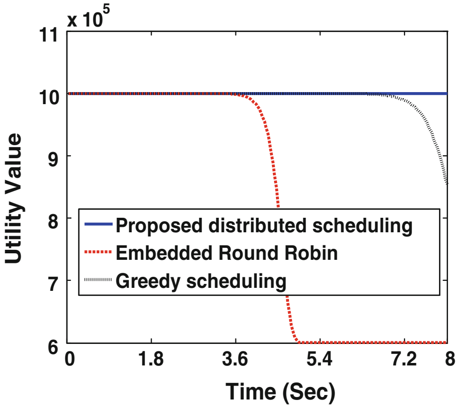
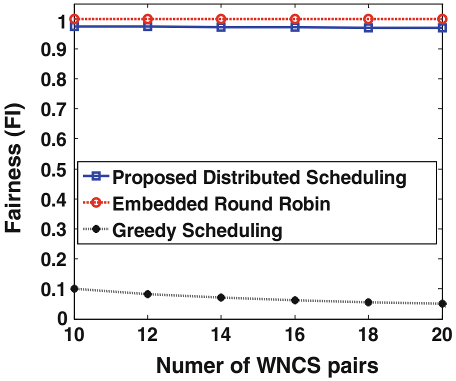


Fig. 8 Fairness comparison



5 Conclusion

In this work through a novel utility-optimal cross-layer co-design, it is demonstrated that the proposed algorithm can optimize not only the performance of the controller, but also the wireless network. The stochastic optimal control does not require system dynamics and wireless network latency and packet losses which are quite useful for hardware implementation, and scheduling algorithm is utility-optimal and distributed which is simpler and requires less computation than centralized scheduling algorithms.

References

1. Branicky MS, Phillips SM, Zhang W, (2000) Stability of networked control systems: explicit analysis of delay. In: Proceedings of the, (2000) American Control Conference. IEEE Press, Chicago, USA, pp 2352–2357:2000
2. Nilsson J, Bernhardsson B, Wittenmark B (1998) Stochastic analysis and control of real-time systems with random delays. Automatica 34:57–64
3. Schenato L, Sinopoli B, Franceschetti M, Polla K, Sastry S (2007) Foundations of control and estimation over lossy networks. In: Proceedings of IEEE, vol 95, pp 163–187, 2007.
4. IEEE Standard 802.11a 1999 (R2003): Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, High-speed Physical Layer in the 5 GHz Band (2003).
5. Dai JG, Prabhakar B (2000) The throughput of data Switchs with and without Speedup. In: Proceedings of 31th IEEE international conference on computer communications, IEEE Press, Tel-Aviv, Israel, pp 556–564, 2000.
6. Vaidy N, Dugar A, Gupta S, Bahl P (2005) Distributed fair scheduling in wireless LAN. IEEE Trans Mob Comput 4:616–629
7. Li Q, Negi R, (2010) Greedy maximal scheduling in wireless networks. In: Proceeding of, (2010) IEEE global communication conference. IEEE Press, Miami, USA, pp 1–5:2010
8. Zheng D, Ge W, Zhang J (2009) Distributed opportunistic scheduling for Ad Hoc networks with random access: an optimal stopping approach. IEEE Trans Inf Theory 55:205–222

9. Bennett JCR, Zhang H (1996) WF2Q: Worst-case fair weighted fair queuing. In: Proceedings of 15th IEEE international joint conference of computer societies and networking of the next generation, IEEE Press, San Francisco, USA, pp 120–128, 1996.
10. Lewis FL, Syrmos VL (1995) Optimal control, 2nd edn. Wiley, New York
11. Xu H, Jagannathan S, (2011) Stochastic optimal control of unknown linear networked control system using Q-learning methodology. In: Proceedings of, (2011) American Control Conference. IEEE Press, San Francisco, pp 2819–2824:2011
12. Jagannathan S (2007) Wireless Ad-hoc and sensor networks: protocols, performance, and control. CRC Press, Florida

On the Estimation Capacity of Equal Gain Diversity Scheme Under Multi-path Fading Channel

Moses Ekpenyong, Joseph Isabona and Imeh Umoren

Abstract This paper investigates analytically-derived close-form expressions for capacity per unit bandwidth in Rayleigh fading channels with equal gain combination (EGC) diversity. We consider power and rate adaptation with constant transmit power, channel inversion bit rate, and truncated channel inversion adaptation policies. To evaluate the performance of the system, we simulate under ideal communication conditions, the different diversity schemes and compare their performance. We discover after extensive simulation that the truncated channel inversion policy outperforms other adaptation policies, while the constant power policy provides the lowest capacity, compared to other policies.

Keywords Diversity schemes · Multi-path fading · Channel gain · Rayleigh fading

1 Introduction

There has been an increasing demand for wireless communication services in recent times. This increase has generally affected the allocation of the scarce radio spectrum, due to an increase in the number of mobile users. In order to maintain the desired level

M. Ekpenyong (✉)
Academic Visitor, School of Informatics, University of Edinburgh,
Edinburgh EH8 9AB, Nigeria
e-mail: ekpenyong_moses@yahoo.com
e-mail: mosesekpenyong@gmail.com

J. Isabona
Department of Basic Sciences, Benson Idahosa University, PMB. 1100,
Benin City, Benin, Nigeria
e-mail: josabone@yahoo.com

I. Umoren
Akwa Ibom State University, Mkpat Enin, Nigeria
e-mail: hollymeh_u@yahoo.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_16,

of productivity and contend with the ever increasing user's capacity, mobile network operators must conserve, share and use bandwidth efficiently. Thus, channel capacity is of fundamental importance in the design of wireless communication systems, as it determines the maximum attainable throughput of the system. A relevant concept is the coherence time [1], which is a measure of the time duration over which channel gain remains almost constant or highly correlated, with a correlation coefficient above 0.5.

Mobile radio links are adversely affected by multi-path fading due to a combination of randomly delayed, reflected, scattered and diffracted signal components. This propagation effect results in time-variation of the signal strength between the transmitter and the receiver. The small-scale fading factor due to multi-path propagation occurs on a time scale of milliseconds (depending on the speed of mobility) which can significantly affect the performance of modulation and coding techniques over the link. This fast fluctuation of the signal envelope is as a result of the convergence of the constructive and destructive multiple reflected radio wave paths. In a dense environment with sufficient scatters (rich scattering), the multi-path attenuation can be effectively modeled by a complex Gaussian random variable at base band. This attenuation is known as *Rayleigh* fading and is used in the evaluation of wireless link performance. The signal fading rate relative to the data rate is also an important parameter.

The end-to-end modeling and design of systems that mitigates the effect of fading are usually more challenging than those whose sole source of performance degradation is the Additive White Gaussian Noise (AWGN). Fading mitigation is important because wireless systems are prone to fading which is also known to cause degradation in the wireless link performance, and calls for efficient fading mitigation schemes. In this paper, closed form expressions for spectral efficiency in a *Rayleigh* fading channel with equal gain combining power diversity and rate adaptation policies are studied and implemented through computer simulations.

2 Review of Related Works

Diversity techniques are becoming well known techniques for combating the notorious effect of channel fading. This is evident in the numerous research works available in the study of channel capacity over fading channels. Initial investigation in this area (use of diversity schemes) dates back to [2], who studied the use of maximum ratio diversity combination (MRDC) technique to provide maximum capacity improvement. In [3], the theoretical spectral efficiency limits of adaptive modulation in Nakagami multi-path fading (NMF) channels are investigated. They apply the general theory developed in [4] to obtain closed-form expressions for the capacity of *Rayliegh* fading channels under different adaptive transmission and diversity combining techniques. In [5], the capacity of a single-user flat fading channel with perfect channel measurement information at the transmitter and the receiver is derived for various adaptive transmission policies. The basic concept of adaptive transmission is the real-time balancing of the link budget through adaptive variation of the transmitted power level, symbol transmission rate, constellation size, coding rate or a combination of these parameters [6].

Mobile radio links are exposed to multi-path fading due to the combination of randomly delayed reflected, scattered and diffracted components. In [7], a novel closed-form expression for achieving average channel capacity of a generalized selection combining rake receiver in Rayleigh fading is derived. A performance comparison of the capacity achieved with maximum ratio combination and rake receivers is also presented. Perera et al. [8] has investigated the limits of information transfer over a fast *Rayleigh* fading Multi Input and Multi Output (MIMO) channel, where neither the transmitter nor the receiver is aware of the channel state information (CSI), except the fading statistics. Their work develops a scalar channel model in the absence of phase information for non-coherent *Rayleigh* fading and derives a capacity supremum with the number of receive antennas at any SNR using La-grange optimization. In [9], a unified L-branch equal gain combining (EGC) over generalized fading channels, such as Nakagmi-m, Rician, Hoyt or Weibull is presented. For each of these models, an exact closed-form expression is derived for the moments of the EGC output and SNR.

In this paper, an extension of the work done in [9] is presented. We derive closed-form expressions for spectral efficiency of *Raleigh* fading channels, with EGC diversity for various adaptation policies. A methodology for computing the optimal cutoff SNR required for successful data transmission is also presented.

2.1 *Equal Gain Diversity Reception for Rayleigh Fading Channel*

Given an average transmit power constraint, the channel capacity of fading channel with received SNR distribution, $P_\gamma(\gamma)$, and power and rate adaptation, C_p bits/s is given as [6]:

$$C_\rho = B\int_{\gamma_0}^{\infty}\log_2\left(\frac{\gamma}{\gamma_0}\right)\rho_\gamma(\gamma)\partial\gamma \tag{1}$$

where B(Hz) is the channel bandwidth and γ_0 is the cutoff level SNR below which data transmission is suspended. This cutoff must satisfy the following condition:

$$\int_{\gamma_0}^{\infty}\left(\frac{1}{\gamma_0}-\frac{1}{\gamma}\right)\rho_\gamma(\gamma)\partial\gamma = 1 \tag{2}$$

where $\rho_\gamma(\gamma)$ represents the *pdf* of the received signal amplitude for a *Rayleigh* fading channel with a m-branch EGC diversity and is given by

$$\rho_\gamma^{EGC}(y) = \frac{y^{m-1}e^{-y/\rho_x}}{\rho_x^m(m-1)!} \tag{3}$$

To achieve the capacity in Eq. (1), the channel fading level must be tracked both at the receiver and transmitter and the transmitter has to adapt its power and rates for excellent channel conditions (i.e., γ is large), by maintaining lower power levels and rates for unfavourable channel conditions (γ is small). Since no data is sent when $\gamma < \gamma_0$, the optimal policy suffers a probability of outage, $\mathrm{P_{out}}$, equivalent to the probability of no transmission, given by:

$$P_{out} = \int_0^{\gamma} \rho_{\gamma}(\gamma)d\gamma = 1 - \int_{\gamma_0}^{\infty} \rho_{\gamma}(\gamma)d\gamma \tag{4}$$

Substituting Eq. (3) into (2), and simplifying same, we observe that γ_0 must satisfy

$$\begin{aligned}\Lambda^{(C)}\left(m, \frac{\gamma}{\rho_x}\right)\frac{-\gamma_0}{\rho_x} \bullet \Lambda^{(C)}\left(m-1, \frac{\gamma_0}{\rho_x}\right)\\ = (m-1)!\frac{\gamma_0}{\rho_x^{m-1}}\end{aligned} \tag{5}$$

where $\Lambda^{(C)}(\alpha, x) = \int_x^{\infty} t^{\alpha-1}e^{-t}\mathrm{dt}$ is the complementary incomplete gamma function [10].

Let $v = \gamma/\rho_x$, and $f(v)$ is defined as:

$$f(v) = \Lambda^{(C)}(m, v) - v\Lambda^{(C)}(m-1, v) - \frac{v}{\rho_x^{m-2}}(m-1)! \tag{6}$$

Substituting Eq. (3) into (1), we have

$$\frac{C_p^{EGC}}{B} = \frac{1}{\rho_x^m(m-1)}\int_{\gamma_0}^{\infty} \log_2\left(\frac{\gamma}{\gamma_0}\right)\gamma^{m-1}e^{-\gamma/\rho_x}d\gamma \tag{7}$$

2.2 Flat Fading and Frequency Selective Fading

As the carrier frequency of a signal is varied, the magnitude of change in amplitude also varies. The coherence bandwidth measures the minimum separation in frequency after which both signals experience uncorrelated fading [11]. In flat fading, the coherence bandwidth of the channel is usually higher than the bandwidth of the signal. Therefore, all frequency components of the signal will experience the same magnitude of fading.

In frequency selective fading, the coherence bandwidth of the channel is lower than the bandwidth of the signal. Different frequency components of the signal therefore experience de-correlated fading, since different frequency components of the signal are independently affected.

3 System Model

3.1 Optimal Rate Adaptation to Channel Fading with Constant Transmit Power Policy

With optimal rate adaptation to channel fading at a constant transmit power, the channel capacity, C_o (bits/s) becomes [4]:

$$C_o = \beta \int_0^\infty \log_2(1+\gamma) P_\gamma(\gamma) d\gamma \tag{8}$$

Substituting Eq. (3) into (8), we obtain

$$\begin{aligned} \frac{C_o^{EGC}}{B} &= \log_2(e) \frac{1}{\rho_x^m (m-1)!} \times \int_0^\infty \log_e(1+\gamma) e^{-\xi\gamma} \gamma^{m-1} d\gamma \\ &= \log_2(e) \frac{1}{\rho_x^m (m-1)!} I_m(\xi) \end{aligned} \tag{9}$$

where

$$I_m(\xi) = \int_0^\infty t^{n-1} \log_e(1+t) e^{-\xi t} dt,\ \xi > 0$$

m represents the diversity levels
ρ_m^x is the average SNR
Using the result of $I_m(\xi)$, we can rewrite Eq. (9) as:

$$\frac{C_o^{EGC}}{B} = \log_2(e) e^{\xi} \Lambda^C(-k, \xi) \tag{10}$$

which can also be expressed in the form of a Poisson distribution as [12]:

$$\frac{C_o^{EGC}}{B} = \log_2(e)(P_M(-\xi) E_1(\xi)) + \sum_{k=1}^{M-1} \frac{P_k(\xi) P_{M-k}(-\xi)}{k} \tag{11}$$

where $P_k(\xi)$ is given by

$$P_k(\xi) = e^{-\xi} \sum_{j=0}^{k-1} \frac{\xi_j}{j!} \quad \text{and} \quad E_1(\xi) = \int_1^\infty \frac{e^{-\xi t}}{t} \text{dt}.$$

3.2 Channel Capacity with Fixed Rate Policy

The channel capacity when the transmitter adapts its power to maintain a constant SNR at the receiver or inverts the channel fading is also investigated in [4]. This technique uses fixed rate modulation and a fixed code sign, since the channel after channel inversion appears as a time invariant AWGN channel. As a result, channel inversion with fixed rate is the least complex technique to implement, assuming good channel estimates are available at the transmitter and receiver. With this technique, the channel capacity of an AWGN channel is given as [4]:

$$C_c = B\log_2\left(1 + \frac{1}{\int_0^{\infty}(\rho_\gamma(\gamma)/\gamma)d\gamma}\right) \tag{12}$$

Inverting channels with fixed rate suffers large capacity penalty relative to other techniques, since a large amount of the transmitted power is required to compensate for the deep channel fades.

The capacity with truncated channel in varied and fixed rate policies, $C_t(bits/s)$ becomes:

$$C_c = B\log_2\left(1 + \frac{1}{\int_{\gamma_0}^{\infty}(\rho_\gamma(\gamma)/\gamma)d\gamma(1 - P_{out})}\right) \tag{13}$$

where $\mathrm{P_{out}}$ is given in Eq. (4). We then obtain the capacity per unit bandwidth with EGC diversity for channel inversion with fixed rate policy (total channel inversion) as C_c/B, by substituting Eq. (3) into Eq. (10) giving:

$$\frac{C_c}{B} = \log_2\left[1 + \frac{\rho_x^m\Gamma(m)}{\int_0^{\infty}\gamma^{m-2}e^{\gamma/\rho_x}d\gamma}\right] \tag{14}$$

Now, substituting $t = \gamma/\rho_x^m \; and \; dt = d\gamma/\rho_x^m$ into Eq. (12), yields,

$$\frac{C_c^{EGC}}{B} = \log_2\left[1 + \frac{\rho_x\Gamma(m)}{\int_0^{\infty}t^{m-2}e^{-t}\mathrm{dt}}\right] = \log_2[1 + (m-1)\rho_x] \tag{15}$$

The capacity of this policy for a *Rayleigh* fading channel is the same as the capacity of an AWGN channel with equivalent $SNR = (m-1)\rho_x$. Truncated channel inversion improves the capacity in Eq. (13) at the expense of the outage probability, P_{out}^{EGC}. The capacity per unit bandwidth of truncated channel inversion with EGC diversity C_t^{EGC}/B, is obtained by substituting Eq. (3) into Eq. (11). Thus,

$$\frac{C_t^{EGC}}{B} = \log_2\left[\frac{1+\rho_x^m (m-1)!}{\int_{\gamma_0}^{\infty} \gamma^{m-2} e^{-\gamma/\rho_x}}\partial\gamma\right] \times (1 - P_{out}^{EGC}) \tag{16}$$

where P_{out}^{EGC} is given in Eq. (3). Substituting $t = \gamma/\rho_x$ and $dt = d\gamma/\rho_x$ into Eq. (14), gives:

$$\frac{C_t^{EGC}}{B} = \log_2\left[1 + \frac{\rho_x (m-1)}{\int_{\gamma_0/\rho_x}^{\infty} t^{m-2} e^{-t} \mathrm{dt}}\right] \bullet \int_{\gamma_0}^{\infty} P_{\gamma}^{EGC}(\gamma)\partial\gamma \tag{17}$$

Substituting $t = \gamma/\rho_x$ and dt $= d\gamma/\rho_x$ into Eq. (15), we arrive at:

$$\begin{aligned} \frac{C_t^{EGC}}{B} &= \frac{1}{\Gamma(m)} \log_2\left[1 + \frac{\rho_x \Gamma(m)}{\Lambda^{(C)}(m-Q,\mu)}\right] \times \int_{\gamma_0/\rho_x}^{\infty} t^{m-1} e^{-t} dt \\ &= \frac{\Lambda^{(c)}(m, \gamma_0/\rho_x)}{(m-1)!} \log_2\left[1 + \frac{\rho_x (m-1)!}{\Lambda^{(c)}(m-Q,\mu)}\right] \end{aligned} \tag{18}$$

4 Simulation and Discussion of Results

To provide real-life perspective of the model design, a simulation of the system was carried out using the three channel capacity schemes proposed as follows:

1. Channel capacity with optimal rate adaptation at constant transmit power
2. Channel capacity with EGC diversity + channel inversion with fixed rate policy (total channel inversion)
3. Channel capacity with EGC diversity + truncated channel inversion.

The simulation was implemented using the matrix laboratory programming toolkit and the results presented in the form of graphs for easy interpretation. The sample inputs were data observed under ideal environmental conditions and is shown in Table 1.

Figures 1, 2 and 3, show the effect of channel capacity on the average SNR for a *Rayleigh* fading channel at different diversity levels (M = 4, 6, 8, 10) for all the transmission policies. We observe that as the diversity levels increases, the channel capacity also increases and improves for all the policies. However, Figs. 2 and 3 show some improvement on the diversity scheme—channel capacity with optimal

Table 1 Sample input parameters and values

Parameter	Value
Diversity levels (m)	4, 6, 8, 10
Average SNR (P_x)	5–50 dB

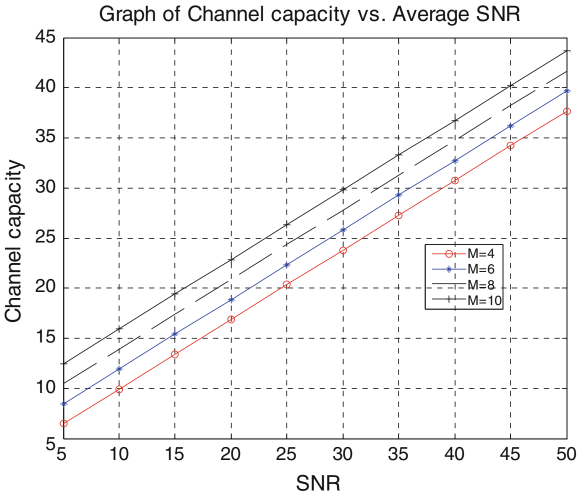


Fig. 1 Graph of channel capacity versus SNR for channel capacity with optimal rate adaptation policy, at constant transmit power

rate adaptation policy at constant transmit power (Fig. 1). These models are capable of remarkably improving the channel capacity, at each diversity level. But, the third model (channel capacity with EGC diversity+truncated channel inversion) presented in Fig. 3, yields the best system performance compared to the other diversity schemes

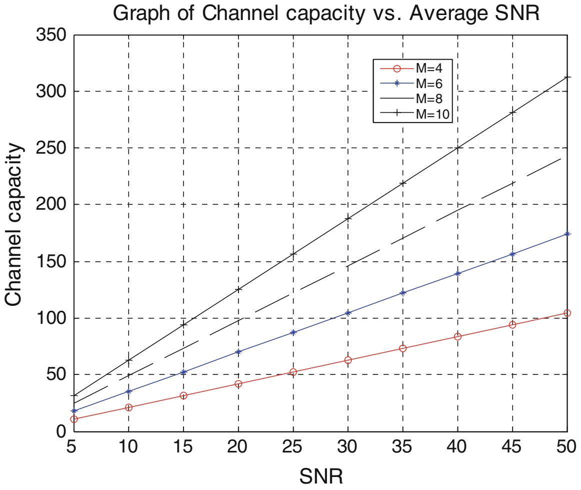


Fig. 2 Graph of channel capacity versus SNR, for channel capacity with EGC diversity and channel inversion with fixed rate policy

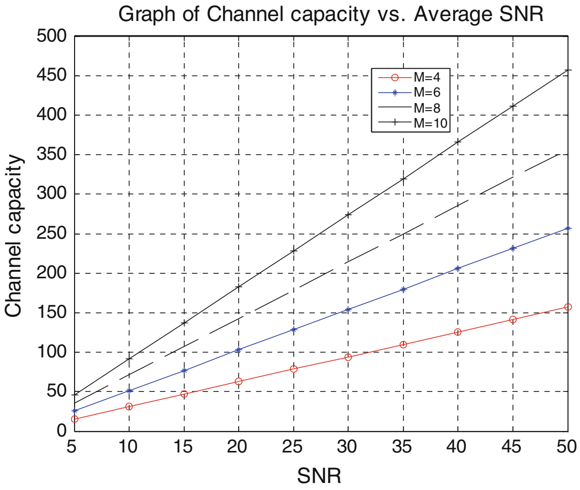


Fig. 3 Graph of channel capacity versus SNR for channel capacity with EGC diversity and truncated channel inversion

(i.e., channel capacity with optimal rate adaptation policy, at constant transmit power and channel capacity with EGC diversity + channel inversion with fixed rate policy).

5 Conclusion

This paper has provided an estimation of the capacity of equal gain diversity schemes under a multi-path fading channel. In order to evaluate these diversity schemes, we simulated the schemes under ideal communication conditions using a robust toolkit. From the simulation results, channel inversion yields the best system performance compared to other diversity schemes. Hence, to maintain the desired level of productivity and contend with the ever increasing users capacity, mobile network operators must conserve, share and manage available bandwidth efficiently.

References

1. Rappaport TS (2002) Wireless communications—principles and practice, 2nd edn. Prentice Hall PTR, Englewood Cliffs
2. Brennan D (1959) Linear diversity combining techniques. Proc IRE 47:1075–1102
3. Alouini M, Goldsmith A (1999) Capacity of Rayleigh fading channels under different adaptive transmission and diversity combining techniques. IEEE Trans Veh Technol 48(4):1165–1181
4. Goldsmith A, Varaiya P (1997) Capacity of fading channels with channel side Information. IEEE Transm Inf Theory 43(6):1986–1992

5. Alouini M, Goldsmith A (1997) Capacity of Nakagami multipath fading channels. In: Proceedings of the IEEE vehicular technology conference VTC'97 Phoemix, A2, pp 358–362
6. Chua S, Goldsmith A (1996) Variables rate variable Power M-QAM for fading channels. In: Proceedings of the IEEE vehicular technology conference, pp 815–819
7. Sagias N, Varzakas P, Tombras G, Karagiannidias G (2004) Average channel capacity for generalized-selection combining RAKE receivers. Euro Trans Telecommun 15:497–500
8. Perera R, Pollock T, Abhayapala T (2005) Non-coherent Rayleigh fading MIMO channels: capacity supremum. In: Proceedings of Asia Pacific conference on communication (APCC), Perth, Australia, pp 72–76
9. Karagiannidis G, Sagias N, Zogas O (2005) Analysis of M-QAM with equal-gain diversity over generalized fading channels. IEEE Proc Commun 152(1):69–74
10. Gradshteyn I, Ryzlink I (1994) Table of integrals, series, and products, 5th edn. Academic Press, San Diego
11. Tse D, Viswanath P (2005) Fundamentals of wireless communication. Cambridge University Press, New York
12. Gunther C (1996) Comment on estimate of channel capacity in Rayleigh fading environment. IEEE Trans Veh Technol 45:401–403

Low Overhead Time Coordinated Checkpointing Algorithm for Mobile Distributed Systems

Jangra Surender, Sejwal Arvind, Kumar Anil and Sangwan Yashwant

Abstract Two checkpointing approaches i.e., coordinated checkpointing and times based checkpointing are widely used in the literature of MDSs. Coordinated checkpointing protocols uses the less checkpoints and domino free but have large coordinated message overheads as processes synchronize by exchanging coordinated message. Time based approach has minimum coordinated messages overheads cost but has high checkpointing cost as it requires large number checkpoints than minimum. Hence coordinated checkpointing approach have minimum checkpointing cost than time based approach but higher coordinated message overheads cost which increases the checkpointing overheads. In this paper, we design an efficient time coordinated checkpointing algorithm which uses time to indirectly coordinate to minimize the number of coordinated message transmitted through the wireless link and reduces the number of checkpoints nearest to the minimum. The algorithm is non-blocking and minimum process.

Keywords Checkpointing · Mobile distributed system · Coordinated · Consistent global state

1 Introduction

Checkpointing/rollback recovery is an attractive and popular technique which gives fault tolerance without additional efforts in DSs [3]. Checkpointing algorithms for DSs have been extensively studied in the literature (e.g., [4, 6, 9]). Due to the emerging challenges of the MDS as low bandwidth, mobility, lack of stable storage, frequent

J. Surender (✉) · K. Anil · S. Yashwant
Department of IT, HCTM Technical Campus, Kaithal 136027, India
e-mail: ssjangra20@rediffmail.com

S. Arvind
Department of CSE, ACE Mithapur (Ambala), Haryana, India

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_17,

disconnections and limited battery life, the fault tolerance technique designed for distributed system cannot directly implemented on mobile distributed systems(MDSs) [1, 4]. A common goal of checkpointing algorithm for MDSs is to reduce the checkpointing cost by taking minimum number of checkpoints and reducing the coordinated messages.

In papers [7–9], authors proposed an efficient time base checkpointing algorithm. The algorithm presented in [9] has lower coordinated message overheads but a global checkpoint consists of all the Nth checkpoints of every process which awoke the processes in doze mode operation. In [7], each process takes its checkpoint at predetermined time instants according to its own local clocks to make the checkpoint consistent. This problem is addressed by using extra messages for clock synchronization. In [8], authors proposed adaptive checkpointing algorithm where they used time to indirectly coordinate the creation of recoverable consistent checkpoints. It requires that checkpoints be sent back only to home agents, which results in high failure-free overhead during checkpointing [2].

Time based protocol requires every process to take checkpoint during checkpointing. Some of the time based protocols defines the timeout period tp. When this timeouts occur, processes take their checkpoint. If the checkpointing interval is too small then multiple checkpoints are transferred from MHs to MSS through wireless link and also checkpointing takes some time to save the application state. This approach increases the checkpointing overheads. On the other hand, if checkpointing interval is too large this may leads to large amount of computational loss during rollback and recovery.

Our algorithm uses time to indirectly coordinate the checkpoint process and avoids checkpointing messages during the checkpoint creation.

2 System Model

The MDS can be considered as consisting of “n” Mobile Hosts (MHs) and “m” Mobile Support Stations (MHSs). All the MSSs are connected through static wired network. A cell is a small geographical area around the MSS supports a MH only within this are and there is a wireless link between a MH to MSS. A MH can communicate to another MH only through their reachable MSS. There are n spatially separated sequential processes denoted by $\mathrm{P}_0, \mathrm{P}_1, \ldots, \mathrm{P}_{\mathrm{n}-1}$, running on MHs or MSSs, constituting a mobile distributed computing system. Each MH/MSS has one process running on it. The processes do not share memory or clock and message passing is the only way for processes to communicate with each other.

As there is no common clock and processes do not share a common memory but every MH and MSS contains a system clock, with typical clock drift rate p in the order of 10^{-5} to 10^{-6}. The system clocks of MSSs can be synchronized using the network time protocol (NTP). MHs start their execution with their own initial timer. Clock can be re-synchronizing with following two methods, to solve the initial time inaccuracies.

3 Algorithmic Concept

Our algorithms is similar to [9], but main disadvantages of the approach is that a global consistent state consists all the Nth checkpoints of every process, where N $\geq$ 0. In this paper we proposed a power efficient checkpointing algorithm that does not awakes to all the MHs; rather only those minimum numbers of MHs that have sent or receives some messages after their last checkpoint will takes checkpoints during checkpointing.

3.1 Initial State

All the process in the system takes their periodically independent from the other processes. Before a checkpointing process starts, a predefined checkpoint period T is set by the local MSS, on the timer of each mobile hosts (MHs) running under it. However, due the varying drift rates of local clock and initial timer inaccuracy the timer at different MHs are not perfectly synchronized. Hence, the checkpoint may not be consistent because of orphan message and there is a great need to resynchronize the local clock. Here we resynchronize the initial or predefined value of T by using the resynchronization mechanism. When the local time of the process expires and if it sends any computation message during this current CI, then it takes checkpoint without the coordination of the other processes, increments in checkpoint sequence number (csn) and reply to its local MSS. On the other hands, if no message has been sent in the current CI then it will takes soft-checkpoint for the current CI and sends reply. Soft checkpoint only shows that process does not send any message during its current CI and its current CI has finished. It has not any transferring cost and stored locally on MHs and discarded when processes takes its checkpoint during next CI. When the MSS knows that it receives checkpoint reply from all the process in minimum set, it removes the previous global state and set new GS or consistent set.

3.2 Selection of Checkpoint Interval

Our algorithm adapts its behavior with the characteristics of the environment. In starting, it will adapt on the basic of quality of service of network. If the network has poor quality of service, the algorithm takes checkpoint frequently else infrequently. The checkpoint period depends upon the quality of service and failure rate of the system. Further, after some time of processing it will learn from the environment and takes decision on the basic of failure rate, average overheads per checkpoints, average latency per checkpoint, and average recovery time as in [9].

3.3 Maintaining Minimum Set

For maintaining the minimum set in proposed checkpointing approach, Each MSS maintain the following information's which are shown in Table 1. Minset [] = (msg_recevdfrom) U (msg_sent_to) Hence, in our approach minset[] is computed by taking the union of column (1) and (2) which contains all the processes from which it received computation message or sent any message, received from the other process.

3.4 Maintaining Consistent Global State

As memory of MHs is not considered stable, when processes take checkpoints, they transferred their checkpoint information including with reply, to its local MSS, where it is stored on the stable storage of MSS. If process does not send any message during its current CI and checkpoint interval expires processes takes soft checkpoint and reply to its local MSS. if MSS receives the reply from all the processes in minset, then a global consistent state is updated. On the other hands if a processes is belongs to minset and does not reply; it sends the reminder and forces it to takes it's soft or permanent checkpoint, and wait for reply.

3.5 Clock Synchronization

MSS is known as the primary server and its timer are used as a reference because it has more reliable timer than those in MHs. In our approach three types of messages are transferred between MH and MSS; computation message, acknowledgement

Table 1 System minimum set record

Mg_rcvdfrom	**Mg_sent_to**	**Mg_status**	**Ack_stat**	**Chkrply_ recvd_from**
(1)	**(2)**	**(3)**	**(4)**	**(5)**
Name of the source process from which msg received)	Name of the destination process to whom recvd msg to be sent	0- msg sent to destination,	0- Ack not sent	Name of the process from which it received checkpoint or soft checkpoint reply.
		1-Ack for sent-msg received	1-Ack sent	
		2- msg sent during previous CI	2-recvd during previous CI	

message and checkpoint reply message. Every application and acknowledgement message is piggybacked with the local clock time of the MSS which is known as $\mathrm{CLK_M}$, and that is, time to next checkpoint. When a process receives the timer with the message, it compares its local time with the one just received. If there are differences in both the timer (received and own), it resets it with the received value. After receiving the piggybacked message, MHs resets its time on the basic of $\mathrm{CLK_M}$.

4 Proposed Checkpointing Algorithm

All the process in the system takes their periodically independent from the other processes. Before a checkpointing process starts, a predefined checkpoint period T is set by the local MSS, on the timer of each mobile hosts (MHs) running under it.

4.1 Data Structure

Each process P_i maintains the following data structure:

In our proposed algorithm we consider a distributed system which has a set of n processes, $\{P_{0,}\ P_1,\ldots,P_{n-1}\}$ where each process $\mathrm{P_i}$ maintains the following data structure.

$\boldsymbol{c_i}$: a Boolean flag $\mathrm{c_i}$ that is initially set at zero. It is set 1only when process $\mathrm{P_i}$ sent a message during current CI, after its latest checkpoints.

csn: checkpoint sequence number of the process which is incremented after taking the checkpoint.

$\boldsymbol{CLK_i}$: clock of process $\mathrm{P_i}$ which shows the time interval until next checkpoint.

$\boldsymbol{CS_i}$: a Boolean flag Checkpoint State $\mathrm{CS_i}$ which is initially set to 0 which shows that process does not takes during its current CI. If process takes checkpoint during its current CI then it will set $\mathrm{CS_i} = 1$.

$\boldsymbol{m_i}$: computation message sent by the process $\mathrm{P_i}$.

$\boldsymbol{SC_i}$: Each process $\mathrm{P_i}$ takes soft checkpoint $\mathrm{SC_i}$ when its local timer expires and does not sends any during its current CI.

$\boldsymbol{Reply_i}$: Each process $\mathrm{P_i}$ sends reply message to its local MSS of taking the permanent and soft checkpoint after the expiries of its timer.

Each MSS maintains the following data structure:

$\mathbf{CLK}_{\boldsymbol{M}}$: clock of the MSS which show the time interval until next checkpoint.

Minset[]: Each MSS maintains the set to minimum number of process which communicate through the MSS.

4.2 An Algorithm

When local clock of process P_i expires

```
If(CLK$_i$ has expired)
  If((c$_i$=1)AND(CS=0))
    {take checkpoint;
    increment in csn;
    set c$_i$=0; //
    continue normal operation ;}
  else if CS=1; //
    set CS=0;
    if SC= T; //
      set SC=F; //
    else
      continue normal operation;
    else
      set SC= T; // take soft checkpoint
```

When P_i receives piggybacked message from P_j when its local clock has not expired

When a process P_i sends computation message by attaching its csn to another process P_j through its local MSS, MSS piggybacked it with time interval to next checkpoint, to the destination process (Fig. 1).

```
Receives message (P$_j$, csn$_j$, CLK$_M$, m$_j$)
  If((csn$_i$ = csn$_j$) AND (CLK$_i$ $\neq$ CLK$_M$))
  reset CLK$_i$ = CLK$_M$;
  receives message
  else if ((csn$_i$ < csn$_j$) AND (c$_i$=1))
  take checkpoint;
  increments csn$_i$;
  set CS$_i$ = 1; set c$_i$ = 0;
  reset timer CLK$_i$ = CLK$_M$;
  process message m$_j$;
  set SC=F;
  else
  resumes normal operation
```

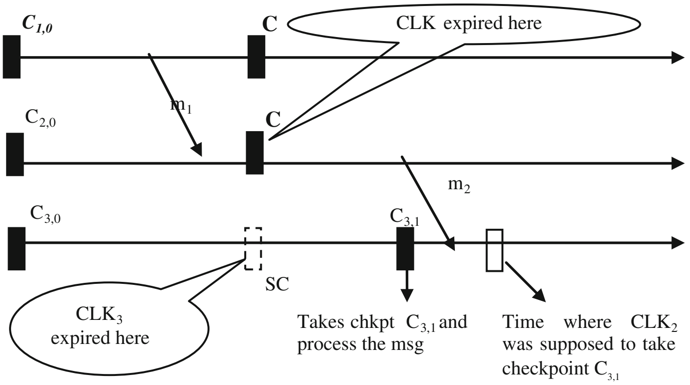

$C_{1,0}$
C
CLK expired here
m_1
$C_{2,0}$
C
m_2
$C_{3,0}$
$C_{3,1}$
SC
CLK_3 expired here
Takes chkpt $C_{3,1}$ and process the msg
Time where CLK_2 was supposed to take checkpoint $C_{3,1}$

Fig. 1 Working of proposed algorithm

4.3 Working Example of Algorithm

A process P_i takes its checkpoint on two ways:

(a) On the expires of its local timer CLK_i: if its timer has expires, then it checks the status of c_i. if $c_i = 1$, then it takes a permanent checkpoint, in another case if $c_i = 0$, it takes soft checkpoint(SC).

(b) On the receipt of piggybacked message in between current CI: in such case it will checks and compares the received csn with its own csn_i. There may be following two possibilities:

(i) msg_csn <= own_csn_i: if it is true, process P_i receives message as a normal without taking any checkpoint. In Fig. 2, csn of process P_1 received with message m_1 is equal to the csn_2. Hence, checkpoint is not taken after receiving the message. However process P_2 takes a checkpoint, after expires of its current CI it sends any message.

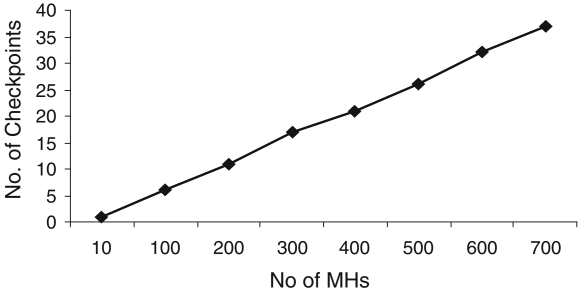


Fig. 2 Number of checkpoints versus Number of MHs

(ii) msg_csn > own_csn_i: it is true, process P_i takes checkpoint first and then precedes the message. If process P_i taken any soft checkpoint, it will discards it and set the checkpoint state (CS) is equal to 1. Furthermore $P_i$does not takes checkpoint, in the current CI, after expires of its local clock CLK_i, if it does not send any message after taking the checkpoint and before expires of the clocks. In Fig. 1, process P_3 takes a checkpoint before receiving the message m_2 as csn received with m_3 is greater than its own csn. Process P_3 receives the message after taking the checkpoint $C_{3,1}$ and it does not takes checkpoint after expires of its local clock.

5 Performance Analysis

In this section we analyze our checkpointing algorithm by comparing with different existing algorithms in different context. We use following notations to compare our algorithm with some of the most notable algorithms in this area of research, namely [4, 6, 8, 9]. A performance comparison is given in Table 2 . In this Table:

C_{air}: Cost of sending a message m from any P_i to P_j;
C_{wired}: Cost of sending a message in the wired link from MSS to MSS.
C_{wl}: Cost of sending a message in the wireless link from MH to MSS.
C_{broad}: Cost of broadcasting a message to all processes;
N_{min}: Number of processes that belong to minset;
N: Total number of MHs in the system;
N_{dep}: Average no. of processes on which a process depends;
T_{ckpt}: Total time taken to store the checkpoint on stable storage

Table 2 compares the performance of our algorithm with the algorithms in [4, 6, 8, 9]. Compared to Naves-Fuchs [8], which is also time-based , our algorithm minimize the half coordinated message including with involving only minimum number of process, so that the total number of checkpoints transmitted onto the wired and wireless network is reduced. Fewer checkpoints and coordinated message transmitted low power consumption for MHs. Compared to [9], our approaches

Table 2 Analytical performance comparison

Algorithm	Blocking time	Checkpoints	Messages
Koo-Toueg [6]	$N_{min} * T_{ckpt}$	N_{min}	$3*N_{min}* N_{dep} * (C_{wired} + C_{wl})$
Cao-Singhal [4]	0	N_{min}	$2*N_{min}*(C_{wired} + C_{wl})$ $+ \min(N_{min}*(C_{wired} + C_{wl}), C_{broad})$
Naves-Fuchs [8]	0	N	$2*N_{min}*(C_{wired} + C_{wl})$
AK [9]	0	N	$N_{min}*(C_{wired} + C_{wl})$
Proposed	0	Nearest to N_{min}	$N_{min} *(C_{wired} + C_{wl})$

involves only minimum number of processes in global state. For the size of the piggybacked information and the coordination message in the wireless link, our algorithm reduces the overheads compared to Cao-Singhal [4] with O(1) to O(N).

Our proposed checkpointing algorithm has the following characteristics:

5.1 Blocking Time

It is clear that the blocking time of our algorithm is 0.

5.2 Power Consumption

In our proposed checkpointing algorithm power efficiently is high compared to [8, 9], as only minimum number of process are involved in determining the consistent global state. It does not awaken the processes in doze mode operation.

5.3 Number of Coordinated Message on Wireless Link

It has very less coordinated message compared to [4, 6] as it is takes decision about their checkpoint independently and only reply message are sent through the wireless links to their local MSS.

5.4 Number of Checkpoints Versus Number of MHs

Our algorithm forces only a minimum number of processes which are directly or transitively dependent on initiator process. Here, we assume in our algorithm as in [5] that 5 % of the MHs of all the MSSs are belong to minimum set and only these participating process receives the checkpoint request message. Figure 2 shows the average number of checkpoints taken by our proposed checkpointing algorithm.

6 Conclusion

In this paper we presented an efficient time based coordinated checkpointing algorithm for mobile computing environments. Our work is an improvement over two phase coordinated checkpointing algorithms [4, 6], and time coordinated algorithms

[8, 9].The algorithm has the following good features which makes it suitable for MDSs:

- It does not use any extra message to coordinate and synchronize the clocks as clocks are attached with the application message which reduces the coordination message overheads.
- It takes reduced number of checkpoints because a process does not take any temporary checkpoint and a process takes checkpoint if and only if it has sent or receives any message during its current checkpoint interval which helps in the efficient use of the limited resources of the mobile computing environment.
- It is non-blocking and takes checkpoint decision independently from the other.
- It does not require tracking and computation dependency information.

Hence our proposed algorithm takes reduced number of checkpoints, minimum interaction (only once) between the MHs and the MSS and no there is no synchronization delay. To achieve these all objective we use very simple data structure. These all features make our algorithm more suitable for mobile computing environment.

References

1. Acharya A, Badrinath BR (1994) Checkpointing distributed application on mobile computers. In: Proceedings of the 3rd international conference on parallel and distributed, information systems, pp 73–80, Sept 1994.
2. Ahn J, Min S, Hwang C (2004) A casual message logging protocol for mobile nodes in mobile computing systems. Futur Gener Comput Syst 20(4):663–686
3. Candy KM, Lamport L (1985) Distributed snapshots: determining global state of distributed systems. ACM Trans Comput Syst 3(1):63–75
4. Cao G, Singhal M (2001) Mutable checkpoints: a new checkpointing approach for mobile computing systems. IEEE Trans Parallel Distrib Syst 12(2):157–172
5. Forman GH, Zahorjan J (1994) The challenges of mobile computing. IEEE Trans Comput 27(4):38–47
6. Koo R, Toueg S (1987) Checkpointing and roll-back recovery for distributed systems. IEEE Trans Softw Eng 13(1):23–31
7. Neogy S, Sinha A, Das P (2002) Distributed checkpointing using synchronized clocks. In: Proceedings of the 26th IEEE annual international conference on computer software and applications (OMPSAC'02), pp 199–206.
8. Neves N, Fuchs WK (1997) Adaptive recovery for mobile environments. ACM Commun 40(1):68–74
9. Singh AK (2007) On mobile checkpointing using index and time together. World Acad Sci Eng Technol 32:144–151

Part III
The Fourth International Conference on Networks & Communications (NETCOM-2012): Measurement and Performance Analysis

Performance Evaluation of TCP NewVegas and TCP Newreno on Burstification in an OBS Network

K. Ratna Pavani and N. Sreenath

Abstract In a TCP over OBS network the TCP data is aggregated at the ingress node and the burst is formed based on time dependent mechanism or quantity dependent mechanism. In time based mechanism, the size of the burst may not be optimal and a small size burst may be formed which may cause wavelength contention to other burst. In quantity based mechanism the bursts are formed with a precise amount of data irrespective of the time needed to form the burst. In case of slow TCP sender and when the data used is non-real-time if we use quantity based mechanism, the TCP source may not generate data since the time of burstification is long and the data generated by TCP is waiting for burstification. This delay in aggregation of data is due to delay in generation of data by the TCP source. As a consequence, burstification cannot be done at a time and the OBS edge node is forced to wait for completion of burstification process. The delay caused due to burstification may influence the performance of TCP variants as it affects the calculation of RTT and therefore size of the congestion window. In this environment we have experimented with two important flavors of TCP, TCP Newreno and TCP NewVegas using NS-2. The results of the simulation prove that TCP NewVegas performs better than TCP Newreno when there is minimum delay due to burstification.

Keywords Transmission control protocol (TCP) · Optical burst switching (OBS) network · TCP-Newreno · TCP-NewVegas · Network simulator version-2 (NS-2)

K. R. Pavani (✉)
Department of Computer Science, Indira Gandhi College of Arts and Science,
Puducherry, India
e-mail: ratnapavanik@gmail.com

N. Sreenath
Department of Computer Science and Engineering, Pondicherry Engineering College,
Puducherry, India
e-mail: nsreenath@pec.edu

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_18,

1 Introduction

There is a tremendous growth in the field of networking over the last few years. As there is an enormous amount of information available at the click of a button, the significance of the Internet is escalating. With the demand for faster and enhanced applications and services, such as WWW browsing, video-on-demand and interactive television, there is an urgent need for the expansion of new high capacity networks that are proficient of supporting these emerging bandwidth requirements [1]. At the same time, we can observe a big advancement in the deployment of broadband access network technologies which place massive traffic on the metro and core transport networks. We need to balance current networks to sustain the escalating volumes of information need. Therefore optical networks are a lucid choice to meet future communication demands. Optical wavelength-division multiplexing (WDM) communication systems have been deployed in various telecommunications backbone networks to meet these rising needs [2]. In these networks bandwidth is divided into number of channels or frequency bands, to ease the access by the end user during peak electronic rates.

Three switching techniques are in existence for WDM all-optical networks, namely optical circuit switching (OCS), optical packet switching (OPS) and optical burst switching (OBS). In an OCS network a dedicated lightpath is established between a source and destination pair [3]. A lightpath is transmitted over a wavelength on each intermediate link. A lightpath must utilize the same wavelength on all the links in the route [4]. This technique is called wave length continuity constraint. For sending and receiving data on the lightpath every node is equipped with transmitters and receivers. During heavy internet traffic setting up a lightpath for long durations causes inefficient utilization of the resources. Apart from this drawback, OCS has one more disadvantage. A lightpath once established may remain alive for days, weeks or months during which there might not be sufficient traffic to utilize the bandwidth.

The OPS concept was evolved to avoid such wastage of bandwidth. It is preferred because optical networks have the potential of switching Internet Protocol (IP) packets directly without converting them into electronics at each intermediate node [5]. In this type of network, an optical packet is sent preceded by its header without any prior path. Using fiber delay lines (FDL), the packet is optically buffered at the core nodes. A switch is configured based on the header information for transmitting the optical packet from input port to the output port. This connection is released as soon as the packets are sent. The practical implementation of these networks demands faster switching times and optical switches that offer switching time of 1–10 ms. In OPS networks that have fixed length packets, packet synchronization becomes indispensable to minimize contention which is complicated to implement [6]. Another major feature of concern in OPS networks is, while using optical buffers like FDLs they are persistently limited by physical space. In order to with hold an optical fiber for a few microseconds, a kilometer of optical fiber will be necessary. Practical implementation of OPS insists on fast switching times, though semiconductor optical amplifiers based switches have lower switching times, they are relatively expensive and optical couplers are used in switch architecture which results in higher power losses [7]. To surmount with the problem of optical buffering and optical processing and to achieve switching in optical domain, OBS networks have been proposed. When compared to OPS and OCS networks, OBS is considered as a balance between the coarse-grained OCS and fine-grained OPS networks [8].

A burst is a basic entity in an OBS network. There are three fundamental components in an OBS, an ingress node, an egress node and a network of core nodes. Ingress nodes and egress nodes can be collectively termed as edge nodes. The edge nodes must collect the IP packets and assemble them into bursts called as burstification. Multiple TCP/IP (transport control protocol/Internet Protocol) connections are aggregated into a data burst at the ingress node to be transmitted into optical domain. In order to transfer a data burst to its destination, a control packet that contains both the burst size and the burst arrival time is created at the network edge node which travels through the core network preceding it [9, 10]. To prevent buffering and processing of the optical data burst at core nodes, a control packet or burst header packet that contains the information about the length and arrival time of the data burst is sent ahead with an offset time. This offset time between the control packet and the data burst is adequate to process the burst header packet and configure the switches at the core nodes to allow the data burst to cut through an all-optical path in the core network. Data burst is disassembled back into IP packets at the egress node. Using one-way signaling mechanism like Tell-and-Go TAG, the data burst cuts through all-optical network following the control packet.

Owing to bufferless nature of OBS networks, and one way signaling scheme, the OBS networks will experience Random Burst Losses (RBL) even at minimum traffic loads [11]. These RBL will be interpreted as network congestion by TCP layer. For example, if a burst that has several packets from a sending TCP is dropped due to contention at minimum traffic loads, the TCP sender times out, which leads to false congestion detection by TCP. This false congestion detection is termed as False Time Out (FTO) [12]. When TCP detects this false congestion, it will initiate congestion control mechanism, which will decrease the size of the CW. The policies used for dealing with change in CW vary for various TCP variants. In this paper a study of two TCP variants namely TCP-Newreno and TCP-NewVegas and their deviation in behavior with respect to a change in congestion window (CW) when burst loss occurs is explored. This paper is organized as follows; a study on TCP variants on OBS is done in Sect. 2. In Sect. 3 the results of simulation are discussed. Conclusion based on simulation results is presented in Sect. 4.

2 TCP Variants on OBS

TCP is a reliable, connection oriented end-to-end protocol. It organizes the packets/ segments based on the information received from upper layers. A TCP packet is encapsulated in an IP datagram which contains TCP header and payload. Usually, TCP sender adjusts the number of segments sent in each round to the buffer size advertized by the TCP receiver to check overflow. This is done by means of a window called CW. To make certain that the data transfer is reliable and to avoid congestion in the network, TCP uses acknowledgements from the receiver for every segment sent. Once the packet is set for transmission, TCP buffers a copy of the packet and sets up the timer for the packet. This timer value is set to Retransmission time out (RTO) which can be used to estimate the round trip time (RTT). If a TCP sender does not get an acknowledgement and if timer expires or if TCP sender receives triple duplicate acknowledgement then TCP understands that the network is in congestion and starts congestion control mechanism. While analyzing congestion control mechanisms [4] done in TCP, it can be classified into the following three categories i.e., loss-based, delay-based and explicit notification-based. In this paper a detailed study was made to examine the performance of loss based TCP variants like TCP-Reno, Newreno which evaluates the congestion in network with the help of packet loss. Also the performance of delay based TCP variants like TCP-Vegas, TCP NewVegas which evaluate the size of CW based on the delay caused due to RTT in an all-optical network.

TCP-Reno is basically a dropping-based TCP Variant. It considers a packet loss as a sign of network congestion and follows an additive increase multiplicative decrease (AIMD) window-based congestion control mechanism [13]. During to a time out in the slow start stage the CW is initialized to the size of one segment and is linearly increased for each successfully acknowledged segment. This linear growth will continue until the current CW is greater than or equal to receiver advertized window or a packet loss event occurs. This results in a TCP sender to enter congestion detection stage. When a time out occurs TCP sender understands that network is heavily congested and at once enters a congestion avoidance stage preceded by a slow start stage. When there is a triple duplicate acknowledgement, TCP sender understands the network to be moderately congested and enters fast retransmission stage by taking half of the sender's CW as the slow start threshold and setting the CW to slow start threshold plus three segments. The TCP sender then tries to retransmit the missing segments and increments the CW by one. When the acknowledgment for the second data segment is received the size of the CW is set to the size of the slow start threshold.

When there are multiple packet losses the performance of TCP-Reno will deteriorate. TCP-Newreno tries to modify TCP-Reno in this aspect [14]. Execution of fast retransmit algorithm of TCP-Newreno is similar to that of TCP-Reno. TCP-Newreno retransmits the lost packet and starts fast recovery phase when there is a triple duplicate acknowledgement. The remaining packets are retransmitted by the sender in as many RTTs as the number of packets in a window, thereby retransmitting one packet per RTT. Through this, performance of TCP-Newreno is hampered by the fact that it takes one RTT to detect a packet loss. The loss of other segments can only be detected when the acknowledgement for the first retransmitted segment is received.

TCP Vegas uses RTT measurement to determine the available network capacity. It does not rely on lost packets like TCP-Newreno to estimate network. Once for every RTT, TCP-Vegas calculate the estimated throughput and actual throughput

[15]. Their difference is then used to evaluate the number of packets that are queued in the network. If the difference exceeds the threshold value then TCP Vegas terminates slow start and commences congestion-avoidance. Unlike TCP Reno and TCP Newreno, TCP Vegas has the ability to terminate slow-start before CW exceeds the network's available capacity. At this stage TCP Vegas decreases the congestion window by one eighth of its current size in order to guarantee that the network does not remain congested. During slow-start phase, for every second RTT CW is increased by one segment per acknowledgment. In congestion-avoidance phase, for every RTT CW will be increased by one segment or decreased by one segment or is left unchanged [16].

TCP NewVegas implements three sender-side changes based on TCP Vegas. Initially, the addition of packet pacing is used to spread the transmission of packets over the entire RTT. This prevents packets from being sent when there is a coarse RTT measurement caused by ephemeral queues. As a second step TCP NewVegas does packet pairing. This is implemented by sending two packets at a time and incrementing the CW during slow-start phase by two segments on every second acknowledgment. By setting the initial window size to an even number, this process guarantees that no packet is sent without having a second one to follow it. Finally, during the congestion-avoidance phase to increase CW more promptly a technique known as rapid window convergence (RWC) is executed. With this technique the size of the CW would grow at a much higher pace till it reaches an optimal value in the network.

In a TCP over OBS network the TCP data is collected by the ingress node and the burst is formed. There are two variations in the formation of burst: time dependent and quantity dependent. In time dependent mechanism the burst is formed with the data collected in stipulated time. Here the burst is transferred without considering the amount of data. In quantity dependent mechanism burst is formed with a limit on amount of data. In this process the burst is formed only after a definite amount of data is collected irrespective of the time needed to form the burst. There also exists another technique which is a combination of these two mechanisms. In that technique burst is either transferred if it meets the required amount of data or if the time taken for burstification process is complete. It may be preferred to send data using quantity based mechanism of burstification when the data is considered to be non real time and TCP source is considered to be very slow. If time based mechanism is used, then the data in the burst will be very low and the burst transmission process may reserve wavelength unnecessarily which may tend to create wavelength contention for some other burst.

Sometimes when we use quantity based mechanism the TCP source may not generate sufficient data and therefore the time of burstification is long. In this situation the data generated by TCP is waiting for burstification and hence the acknowledgement may not be received by the TCP. Due to this the RTT at the source may time-out. Source TCP may assume this situation as congestion in network and react to it by lowering the CW size.

This problem is similar to multiple packet loss issue in Newreno or the way the congestion is dealt in NewVegas. In TCP Newreno, when there is multiple packet loss it takes one RTT per packet to recover from slow start phase. And in case of TCP NewVegas which uses RWC technique to increase its CW size to an optimal value and therefore recovers from slow start phase straight away. Hence it is presumed that TCP NewVegas with its fast recovery from slow start phase after network congestion will have an enhanced performance over TCP Newreno in an OBS network. In this backdrop the two TCP variants, TCP-Newreno and TCP-NewVegas are simulated using Network simulator version 2.27 (NS-2.27) to evaluate their performance. The environment under which this simulation is done has been explained here. An OBS patch [17] is put on NS-2.27. Variants are simulated using random uniform burst distribution algorithm. NSFNet topology with 14 core nodes, 28 TCP nodes and 10 TCP/IP connections is designed to perform the simulation. In OBS network, in the edge nodes packets are aggregated into burst and transmitted all optically from source to destination. Optical classifier in the OBS patch processes the packets in the core network. The packets that are assembled in a single burst are defined in Burst size and this process is called Burstification.

The performance of TCP variants is evaluated by varying the size of the burst and by changing the duration of burstification. In this simulation the size of the burst varies from 10 to 11000 packets per burst. The delay between the control packet and the burst is set to 0.01 ms. The assembly buffers at the ingress node are set to 100. Period of burstification is varied between 0.001 and 1.0 to estimate its impact on throughput/burst-delivery-ratio of TCP variants.

The data plane and the control plane are separated in OBS simulator to eliminate the problems in all-optical processing of packet headers. JET signaling mechanism is used the core network. The control packet has all the essential information so that each intermediary optical switch in the core OBS network can transmit the data burst and also configures its switching matrix in order to switch the burst all-optically. In the core OBS network the conversion of electrical-optical-electrical is taken care by the edge node. Control packets are generated and forwarded followed by the data burst in the core network. TCP segments from the optical bursts are separated at the node entrance by a classifier. In this experiment the routing and scheduling in core OBS network is done by latest available unused channel with void filling (LAUC-VF) [18] and minimum starting void (Min-SV) [19] algorithms.

3 Results

In Fig. 1 the topology that used for the simulation is shown. In this Network there are 14 optical core nodes and 28 electrical nodes. The core network in this simulation is modeled as a single network with 1 Gbps bandwidth and 10 ms propagation delay. There is a bandwidth of 155 Mbps with link propagation delay of 1 ms to the access links. In this simulation the burst size varies from 10 to 11000 packets per burst.

Table 1 describes the topology and values used to simulate the TCP variants in an OBS environment using NS-2. Figure 2 shows the simulation results of the two TCP variants, TCP Newreno and TCP NewVegas. It is a graph showing the variation

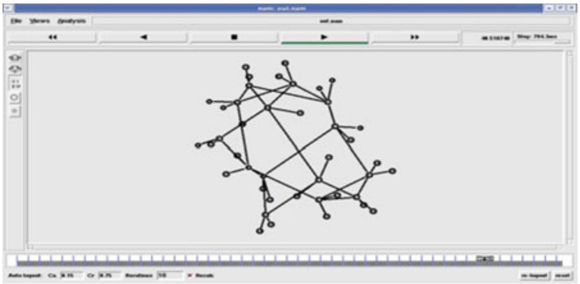
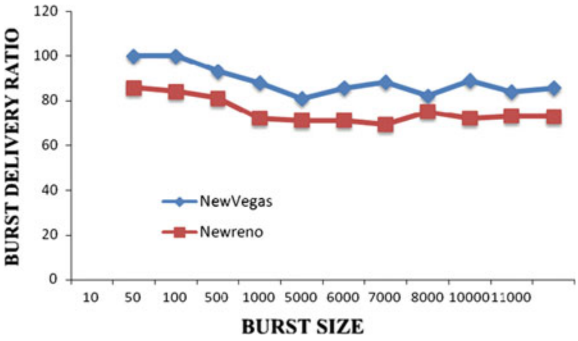


Fig. 1 NSF topology with 14 optical core nodes and 28 electric nodes

Table 1 Simulation parameters

Topology used	NSFNet
Total number of optical core nodes	14
Total number of electronic nodes	28
Total number of TCP/IP connection	10
Packets per burst varies from	10 to 11000
Max lambda value	20
Maximum link speed	1 GB
Hop-delay	0.01 ms
Burstification period value (ms)	0.001, 0.01 and 1.0

Fig. 2 Performance comparison when BTO = 0.001



in performance of the variants when there is a change in the burst-time-out (BTO) value. It is observed that when BTO is 0.001 ms TCP NewVegas performs better than TCP Newreno even when the size of burst is 10. After 7,000 it is observed that there is a decline in performance of TCP Newreno when compared to TCP NewVegas.

In Fig. 3 the performance of the two variants is evaluated when the BTO value is 0.01. When the delay in burstification is reduced, it can be observed from the result that the performance of TCP Newreno in considerably lower than TCP NewVegas (Fig. 4). But when BTO time is increased to 1.0 ms the performance of both the

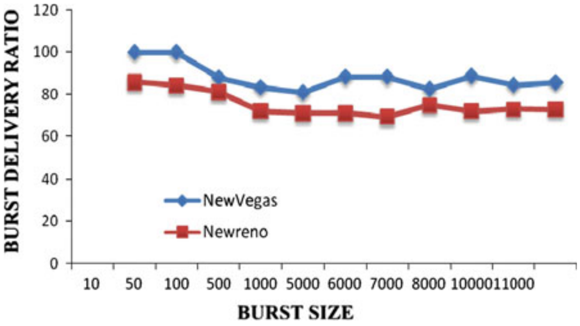


Fig. 3 Performance comparison when BTO = 0.01

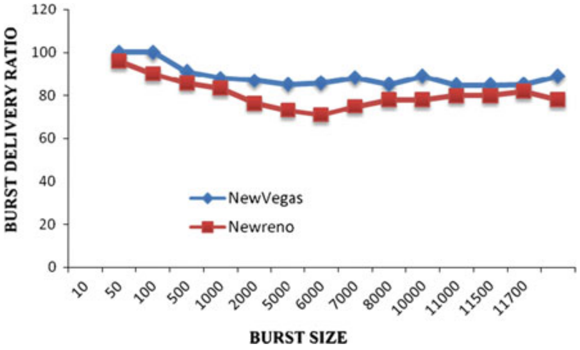


Fig. 4 Performance comparison when BTO = 1.0

variants is almost equal, with TCP NewVegas performing slightly better than TCP Newreno when burst values are above 5000. A study is made on TCP Newreno with BTO = 0.001 and 1.0 to appraise its performance. It can be seen in Fig. 5 that burst delivery ratio of TCP Newreno is higher with BT0 value 1.0 than with BTO 0.001.

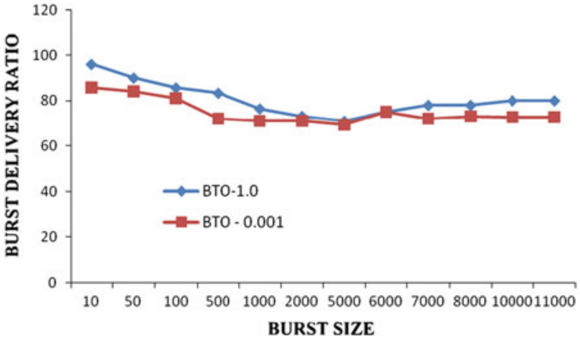


Fig. 5 Performance comparison of TCP-Newreno when BTO is 1.0 and 0.001

4 Conclusion

From the Graphs above it is clearly seen that TCP NewVegas performs much better that TCP Newreno in an all-optical burst switching network. Simulations are done with varying burst size and by altering the BTO values. When the BTO is less than or equal to 0.01 the performance of TCP NewVegas is better than TCP Newreno. In an optical network with higher bandwidth and faster switching time, where a data burst travels cutting through the all-optical network, higher BTO will influence the RTT calculation and thereby the CW. Hence it is always advantageous to have a lower BTO so that there will not be a delay in the TCP source due to burstification. In this scenario, it is observed that TCP NewVegas is more suitable to optical networks as it offers a superior performance over loss based TCP variants.

References

1. Venkatesh T, SivaRamMurthy C (2010) An analytical approach to optical burst switched networks. Springer, New York e-ISBN 978-4419-509-2
2. Chatterjee S, Pawlowski S (1999) All-optical networks. Commun ACM 42(6):74–83
3. Venkatesh T (2009) Estimation, classification, and analysis of losses in optical burst switching networks. A thesis submitted for the award of the degree of Doctor of philosophy, July 2009
4. Siva Ram Murthy C, Mohan G (2002) WDM optical networks: concepts, design, and algorithms. Prentice Hall PTR, New Jersey
5. Yao S, Mukherjee B, Dixit S (2000) Advances in photonic packet switching: an overview. IEEE Commun Mag 38(2):84–94
6. Blumenthal DJ, Prucnal PR, Sauer JR (1994) Photonic packet switches: architectures and experimental implementation. Proc IEEE 92(11):1650–1667
7. Ramaswam R, Sivarajan KN (1998) Optical networks a practical perspective. Academic Press, San Diego
8. Vokkarane VM, Jue JP (2004) Introduction to optical burst switching. Springer, New York
9. Xiong Y, Vandenhoute M, Cankaya H (2000) Control architecture in optical burst switched WDM networks. IEEE J Sel Areas Commun 18(10):1838–1851
10. Yu X, Chen Y, Qiao C (2002) Study of traffic statistics of assembled bursts in optical burst switched networks. In: Proceedings of opticomm. Boston
11. Zhang Q, Vokkarane VM, Wang Y, Jue JP (2005) Analysis of TCP over optical burst-switched networks with burst retransmission. In: Proceedings of IEEE GLOBECOM, St. Louis, MO
12. Yu X, Qiao C, Liu Y (2004) TCP implementations and false time out detection in OBS networks. In: Proceedings of IEEE INFOCOM. Hong Kong, pp 358–366
13. Stevens W (1997) TCP slow start, congestion avoidance, fast retransmit, and fast recovery algorithms, RFC 2001
14. Floyd S, Henderson T, Gurtov A (2004) The NewReno modification to TCP's fast recovery algorithm. Network Working Group, RFC 3782
15. Sing J, Soh B (2005) TCP New Vegas: improving the performance of TCP vegas over high latencyLinks. In: Proceedings of the 2005 fourth IEEE international symposium on network computing and applications (NCA'05). IEEE Computer Society, Washington
16. Sing J, Soh B (2006) TCP New Vegas: performance evaluation and validation. In: Proceedings of the 11th IEEE symposium on computers and, communications (ISCC'06)
17. Gurel G, Alparslan O, Karasan E (2007) nOBS: an ns2 based simulation tool for performance evaluation of TCP traffic in OBS networks. Ann Telecommun 62(5—-6):618–632

18. Xiong Y, Vandenhoute M, Cankaya HC (2000) Control architecture in optical burst-switched WDM networks. IEEE J Sel Areas Commun 18(10):1838–1851
19. Xu J, Qiao C, Li J, Xu G (2003) Efficient channel scheduling algorithms in optical burst switched networks. In: Proceedings of IEEE Infocom'03, 3:2268–2278, 30 March–3 April 2003

Performance Enhancement Through Optimization in FPGA Synthesis: Constraint Specific Approach

R. Uma and P. Dhavachelvan

Abstract Circuit synthesis denotes the automated generation of logic networks from behavioral descriptions at an arbitrary level. Synthesis is becoming a key issue in VLSI design for efficient and flexible usage of cell and component. The architectural synthesis involves resource allocation, resource binding, and scheduling tasks. As the capacity of FPGAs increases, synthesis tools and efficient synthesis methods for targeted device become more significant to efficiently exploit the resources and logic capacity. This paper explores a design solution and synthesis optimization constraints for targeted FPGA device. The issue focuses on: the synthesis optimization for various modeling approaches; synthesizer producing sub optimal results for setting the target constraints too high; how an inefficient coding style can adversely impact synthesis and simulation, resulting in slow circuits. All the design solutions are elucidated with appropriate example. The module functionality are described using Verilog HDL and performance issues like slice utilized, simulation time, percentage of logic utilization, level of logic are analyzed at 90 nm process technology using SPARTAN6 XC6SLX150 XILINX ISE12.1 tool.

Keywords Synthesis target · FPGA · High target constraints · Slice utilization · Data-flow · Behavioral modeling approach

R. Uma (✉) · P. Dhavachelvan
Department of Computer Science, School of Engineering, Pondicherry University,
Puducherry, India
e-mail: uma.ramadass1@gmail.com

P. Dhavachelvan
e-mail: dhavachelvan@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_19,

1 Introduction

The Field-Programmable Gate Array (FPGA) is a new ASIC medium that affords instant manufacturing turnaround and exceptionally low manufacturing costs. FPGA typically incorporates Look up tables (LUTS), slices, I/O buffers, Block RAM (BRAM), number of other hard cores like digital clock managers, systolic filters for DSP core and Adder/ multiplier blocks.

Normally in FPGAs the functional description are specified using HDL. Synthesis tools that automatically map a design composed of simple gates and netlist. As the capacity of FPGAs increases, synthesis tools and efficient synthesis methods for targeted device become more significant to efficiently exploit the resources and logic capacity. Synthesis Optimization is a complex series of transformations guided by user-defined constraints. This tool provides a variety of design constraints which essentially helps the designer to meet the design goal such as area and speed optimization to obtain the best implementation logic.

The synthesis tools optimize HDL code for both logic utilization and performance of an intended design [1, 2]. In FPGA each slices, LUT and register utilization are very vital in order to accommodate larger design unit. This paper explores a design solution and synthesis optimization constraints for targeted FPGA device. The issue focuses on: the synthesis optimization for various modeling approaches; synthesizer producing sub optimal results for setting the target constraints too high; how an inefficient coding style can adversely impact synthesis and simulation, resulting in slow circuits [3].

The organization of the paper is as follows: In Sect. 2, the optimization for various modeling approaches and its performance issues. Section 3 focus the circuit design targeted for high constraints. Section 4 presents the adverse effect of synthesizer and simulator for inefficient coding. Section 5 provides the discussion for these three issues. Finally the conclusion is presented in Sect. 6.

2 Synthesis Optimization: Various Modeling Approach

FPGA synthesis tool provides a variety of design constraints which essentially helps the designer to meet the design goal such as area and speed optimization to obtain the best implementation logic. The following functional coding explains the concept of optimization between different modeling approaches. This code illustrates the design of simple arithmetic logical unit (ALU) which performs the basic operations of addition, subtraction, multiplication and division. This design has been coded with data flow using assignment statement and behavioral modeling using case statement. The design of ALU in data flow modeling is designed with tristate logic. The functional coding is depicted below in Fig. 1. Generally for modeling circuit using case statement the synthesis tool creates a balanced logic and evaluates the expression using common controlling expression. When writing case statements, make sure all

```
Data flow modeling ALU
module ALUdataflow (a, b, s, out);
input [4:1] a,b;
input  [1:0] s;
output [4:1] out;
assign out = s==2'b00 ? a+b :4'bz;
assign out = s==2'b01 ? a-b :4'bz;
assign out = s==2'b10 ? a*b :4'bz;
assign out = s==2'b11 ? a&b :4'bz;
endmodule
```

```
Behavioral modeling ALU
module caseALU (a, b, s, o);
  input [4:1] a,b;   input  [1:0]  s;
  output [4:1] o;    reg    [4:1] o;
  always @(a or b or s)
  begin
    case (s)
      2'b00   : o = a+b;   2'b01   : o = a-b;
      2'b10   : o = a*b;   default : o = a&b;
    endcase
  end
endmodule
```

Fig. 1 Functional coding for ALU design

outputs are defined in all branches. It is desirable to use case statement with "full and parallel" construction in order to avoid latches instead of multiplexers. The design has been synthesized with various options to optimize the design. The optimization goal is set to speed and area by setting optimization effort high, keeping hierarchy soft and encoding option in auto mode. The design synthesizes result of ALU design with dataflow and behavioral modeling is shown in Table 1. When the design of ALU in dataflow modeling is optimized for speed the number of slices utilization is large when compare to the design of ALU in behavioral modeling. This is due to the fact that the **synthesizer procedures balanced logic while synthesizing the code written in case statement than code written using assignment statement**. The case statement converges fast with utilizing fewer resources and also the case statement is full and parallel thereby reduces the formation of latches. Therefore synthezing the design of ALU using case statement with the optimization for area and speed will have the same number slice utilization, delay and logical depth [4].

While synthesizing the dataflow modeling for the optimization constraint for speed the design utilizes more number of slices since the output signal of this design has been tristated. So for this particular design the output signal for the width of 4 uses 16 tristate buffer for 4 different functions like addition, subtraction, multiplication and logical AND operation. Because of this tristated output the slice utilization are maximum and the synthesizer will not produce balanced logic for tristate conditional circuit. If the same design synthesized for the optimization constraint for area the resource utilization is comparatively less when compare to speed optimization and the path delay is also less when compare to behavioral modeling. The comparison of speed and area optimization for ALU design with dataflow and behavioral modeling is shown in Fig. 2. From speed optimization it is observed that the ALU design with dataflow occupied much slices and delay when compare to behavioral modeling. From area optimization considerable reduction in slice usage has been observed for ALU design with dataflow, and delay is less when compare to behavioral modeling.

Table 1 Synthesis of ALU design using dataflow and behavioral modeling

Module name	Constraint setting	Slices utilized		Delay			Levels of logic
		No of slice LUT	No of bonded IOBs	Routing	Logic	Total delay	
ALU design in data flow modeling	Optimization goal: speed Optimization effort: high Keep hierarchy : soft Encoding option : Auto	22	14	3.480 41.8 %	4.850 58.2 %	8.33	6
	Optimization goal: Area Optimization effort: high Keep hierarchy : soft Encoding option : Auto	18	14	2.996 38.1 %	4.868 61.9 %	7.864	6
ALU design in behavioral modeling	Optimization goal: speed Optimization effort: high Keep hierarchy : soft Encoding option : Auto	13	14	3.282 39.2 %	5.029 60.5 %	8.311	7
	Optimization goal: Area Optimization effort: high Keep hierarchy : soft Encoding option : Auto	13	14	3.282 39.2 %	5.029 60.5 %	8.311	7

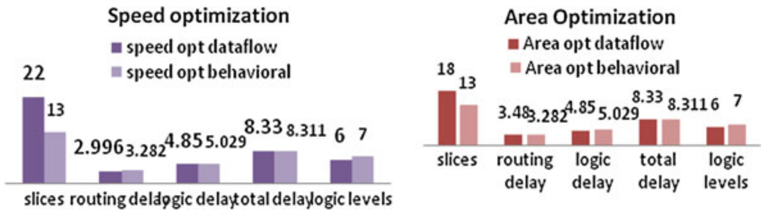


Fig. 2 Speed and area optimization

To consolidate this analysis circuit designed for conditional logic to meet high speed then data flow modeling with area optimization is suitable and for less area behavioral modeling with area optimization is preferable. Design to meet both constraint model the circuit using behavioral style especially design the circuit using case statement rather than if statement. Since if statement creates priority encoding logic therefore utilizes large slices when compare to case statement.

The RTL view of this synthesized design is shown in Fig. 3. For ALU designed with data flow utilizes the internal primitives of adder, subtractor and multiplier unit. The output of this design uses 4 tristate buffer named BUFT. For ALU designed with behavioral employ the internal primitives of adder/subtractor block and multiplier block. The output of this design uses two multiplexers through which one of logic function can be selected [5–7]. The overall analysis provides **designing the module for conditional based logic behavioral modeling is suitable because the synthesizers converges the circuit faster and take less resource to provide balance between the slice utilization and logical depth**.

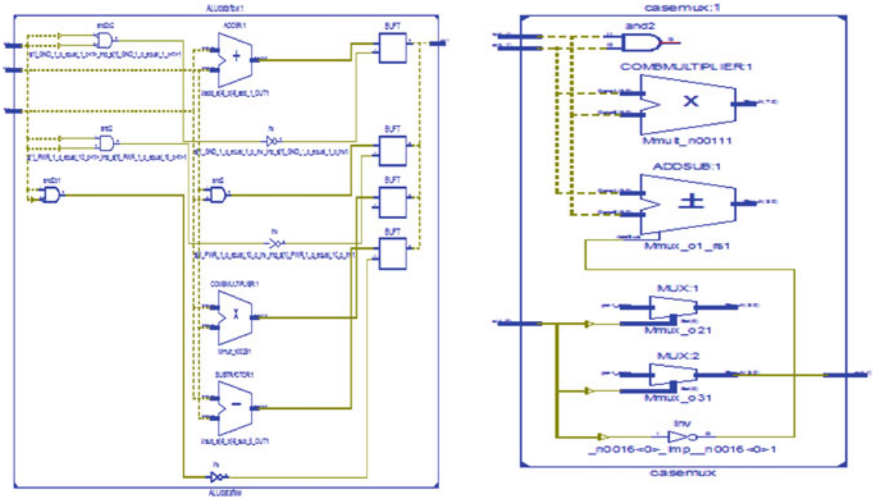


Fig. 3 RTL view for ALU design using dataflow and behavioral modeling

3 Synthesis Optimization: High Target Constraints

This topic presents the impact of the design when the constraints are set too high. The synthesis based gate-level optimizations will include constraints like Finite State Encoding (FSM) algorithm (like auto, one-hot, compact, sequential, gray, Johnson, speed1), hierarchy setting (which allows MAP's physical synthesis options to optimize across hierarchical boundaries), logic duplication (avoids replication of logic), FSM style, register duplication and so on. By setting the required constraints the design can be optimized. As a general rule faster design requires parallelism at the expense of slice area, and minimize area design requires less logical depth. The following example helps to illustrate the design problem by oversetting the constraints. This example depicts the implementation of ripple carry adder initially set with the constraints of speed with no hierarchy and FSM is set in auto and its performance is analyzed. Next the same example is analyzed by setting the constraints of speed, hierarchy in soft mode, high optimization effort, FSM in one-hot mode. It is observed that **if the target constraints are set too high the design can be implemented in a suboptimal manner that actually results in a lower speed, higher slice density and logic depth**. Table 2 presents the circuit synthesized for speed/area with no hierarchy and auto FSM mode. Table 3 presents the synthesis report of the circuit for speed/area with soft hierarchy, high optimization effort and one-hot FSM mode. The functional code for this combinational ripple carry adder is shown in Fig. 4.

When optimizing this circuit with the constraint of speed/area with no hierarchy and normal optimization effort the circuit produces minimum delay, logical levels and slice utilization. This is because in FPGAs the internal architecture for ripple carry adders are optimized already. Therefore when over setting the constraint the circuit produces sub optimal results. **For high constraints the synthesizer produces maximum delay, slice utilization and logical levels. Therefore care should be taken while selecting the constraints to produce better and efficient optimized output**.

Table 2 Speed/Area optimization with no hierarchy, optimization effort is normal and auto FSM mode

Bit size	Delay (ns)	Slices utilized		Logical levels
		LUT slice	Bonded I/Os	
8	8.851	12	26	6
16	13.829	25	50	11
32	21.903	49	98	19
64	38.051	97	194	358

Table 3 Speed/Area optimization with soft hierarchy, optimization effort is high and one-hot FSM mode

Bit size	Delay (ns)	Slices utilized		Logical levels
		LUT slice	Bonded I/Os	
8	12.429	16	26	20
16	20.231	32	50	34
32	35.835	64	98	66
64	67.043	128	194	130

```
module rca_8bit(a,b,d,s,cout);
input [7:0]a,  input [7:0]b, input d, output [7:0]s, output cout;
rca_4bit y1(a[3:0],b[3:0],d,s[3:0],x);
rca_4bit y2(a[7:4],b[7:4],x,s[7:4],cout);endmodule
```

Fig. 4 Functional coding for ripple carry adder design

4 Synthesis Optimization: In-efficient Structural Coding

This section presents how a different modeling approach affects the performance of larger circuit. To elucidate this concept a full adder is implemented with data flow and behavioral modeling. Both the design produces the same results. Full adder named FA uses the library operators like XOR, AND and OR to produce the sum and carry output. The module FA1 is implemented using the truth table of full adder circuit and it is sequential design. The synthesized report of FA and FA1 along with its functional coding is presented in Table 4 and Fig. 5. From the synthesized output it can be observed that the FA has maximum delay of 5.502 ns with maximum logical levels. The FA1 has minimum logical level and delay of 3.381 ns. From this observation it is noticed that the sequential circuit in behavioral modeling synthesis the code faster when compare to dataflow modeling. These differences are due to

Table 4 Simulation and synthesis result of Adder1 and Adder2

Module name	Delay (ns)	Slices utilized	I/O buffers	% Logic	% Routing	CPU to Xst speed (secs)	Logic levels
FA	5.502	2	5	4.178 ns (74.6 %)	1.424 ns (25.4 %)	3.03	3
FA1	3.381	8	5	1.734 ns (49.7 %)	1.757 ns (50.3 %)	2.06	1

```
module FA( a,b,c,s,cout);
input a;  input b; input c;
output s; output cout;
assign s=a^b^c;
assign cout=a&b|a&c|b&c;
endmodule
```

```
module FA1(a,b,c,s,cout);
input a; input b; input c;output s; output cout;
reg s,cout;
always @( a or b or c)
begin
  case ({a,b,c})
   000: begin  s = 1'b0 ;cout=1'b0;end
   001|010|100:begin  s=1'b1;cout=1'b0;end
   011|110|101:begin  s=1'b0;cout=1'b1; end
   111: begin  s = 1'b1 ;cout=1'b1; end
  endcase
  end
endmodule
```

Fig. 5 Functional coding for ALU design

Table 5 Performance of array multipliers with FA, FA1 and core generator

Performance of array multiplier								
Module name	Bit sizes	Delay (ns)	Slices utilized	I/O buffers	% Logic % (ns)	% Routing	CPU to Xst speed (secs)	Logic levels
Array multiplier with FA	4	7.871	4	8	4.918 62.5 %	2.952 37.5 %	3.38	5
	8	10.995	10	16	8.837 86.1 %	2.158 14.9 %	3.46	10
	16	15.21	35	32	12.837 88.1 %	3.258 10.2 %	3.57	20
	32	26.391	1-dsp	64	24.098 89.6 %	2.293 9.4 %	4.01	25
	64	32.31	2-dsp	128	31.123 91.12 %	1.234 9.01 %	4.23	30
Array multiplier with FA1	4	5.103	11	8	2.351 70.2 %	2.351 29.8 %	3.3	3
	8	7.571	1-dsp	16	5.918 63.5 %	2.952 36.5 %	3.12	7
	16	7.995	2-dsp	32	7.837 87.1 %	1.158 12.9 %	3.18	12
	32	9.31	4-dsp	64	9.001 88.1 %	1.258 11.2 %	3.31	16
	64	20.278	8-dsp	128	18.098 84.6 %	3.293 15.4 %	3.45	19
Array multiplier with multiplier core generator	4	5.703	4	8	4.197 73.6 %	1.506 26.4 %	3.12	4
	8	7.871	17	16	4.918 62.5 %	2.952 37.5 %	3.18	8
	16	8.995	1-dsp	32	7.837 87.1 %	1.158 12.9 %	3.31	13
	32	10.21	2-dsp	64	9.837 88.1 %	1.258 11.2 %	3.45	17
	64	21.391	4-dsp	128	18.098 84.6 %	3.293 15.4 %	3.55	20

the number of events that is presented by the design and the simulators. Until recent years, Verilog-XL could simulate data-flow designs much faster than it could simulate behavioral/RTL code.

The adder designed in data and behavioral modeling is used to design larger circuit like array multiplier with bit length of 4, 8, 16, 32 and 64 and its performance comparison is observed. This larger unit is constructed with FA, FA1 and core generator as a component to this wider design. The performance of array multiplier is shown in Table 5.

From this performance analysis it is observed that array multiplier implemented with FA has least device utilization, maximum logical levels and delay. Array multiplier with FA1 has least delay, logical depth and maximum device utilization. Multiplier with core generator produces optimum results in terms of slice utilization and logical depth when compare to FA1. **To conclude this analysis for smaller array multipliers use the structure of FA. If the width size is beyond 16 then use FA1. For larger width more than 32 it is preferable to utilize the core generators**.

5 Discussion

This paper explores the several optimizations in order to meet the desired constraints. The focus has been to minimize the active area, speed and resources in a FPGA target device. The contributions of this paper are summarized below:

Various Modeling Approaches:

- The first issue presents the synthesis optimization for various modeling approach. Circuit designed for conditional logic to meet high speed then data flow modeling with area optimization is suitable and for less area behavioral modeling with speed optimization is preferable.

High Target Constraint:

- The second issue focuses the design of ripple carry adder producing suboptimal results for high target; this is because in FPGAs the internal architecture for ripple carry adders are optimized already. Therefore it is prudent not to set too high constraints for already optimized module.

In-efficient Coding:

- The third issue focuses how in-efficient coding adversely affects the performance of larger device. Because coding style and mega function implementation can have such a large effect on the design performance, it is important to match the coding style to the device architecture from the very beginning of the design process.

6 Conclusion

This paper presents the synthesis optimization for various constraints to minimize the resource utilization and logic density to accommodate the larger design in a FPGA device. Key aspects of this work are: synthesis optimization for various modeling approach to exploit the efficient usage of resource and device logic utilization for various modeling approach for high speed and less area constraints.

References

1. Wang JX, Loo SM (2010) Case study of finite resource optimization in FPGA using genetic algorithm. IJCA 17(2):95–101
2. Sangiovanni-Vincentelli A, El Gamal A, Rose J (1993) Synthesis methods for field programmable gate arrays. Proc IEEE 81(7):1057–1083
3. French M, Wang L, Anderson T, Wirthlin M (2005) Post synthesis level power modeling of FPGAs. In: Proceedings of the 13th annual IEEE symposium on field-programmable custom computing machines (FCCM'05), IEEE, Washington
4. Cassel M, Kastensmidt FL (2006) Evaluating one-hot encoding finite state machines for SEU reliability in SRAM-based FPGAs. In: Proceedings of the 12th IEEE international on-line testing, Symposium (IOLTS'06). IEEE, Washington, pp 145–150
5. Ferrandi F, Lanzi PL, Palermo G, Pilato C, Sciuto D, Tumeo A, Politecnico di Milano (2007) An evolutionary approach to area-time optimization of FPGA designs, IEEE
6. St anislaw Deni ziak, Mar iu s z WiGniews ki (2009) A symbolic RTL synthesis for LUT—based FPGAs, IEEE
7. Cong J, Liu B, Neuendorffer S, Noguera J, Vissers K, Zhang Z (2011) High Level Synthesis for FPGAs: from prototyping to deployment. IEEE Trans Comput Aided Design Int Circuits Sys 30(4):473–491

Performance Optimization of Vehicular Ad Hoc Network (VANET) Using Clustering Approach

Ankita Anand and Parminder Singh

Abstract Vehicular ad hoc networks (VANETS) have actually attracted a lot of attention over the last few years as being used to improve road safety. In this paper, cluster based technique has been introduced in VANET. As VANET is a new form of MANET, so with this cluster based technique in VANET, several handoff problems have been removed, which were actually difficult to remove in MANET. For this traffic infrastructure cluster based routing has been used, with two routing protocols i.e. AODV and AODV+. The network simulator NS2 has been used for removing unpredictable movements that may arise in the network.

Keywords VANET · Clustering · AODV · AODV+

1 Introduction

IEEE 802.11 p is an approved amendment to the IEEE 802.11 standard to add wireless access in vehicular environments [1]. The IEEE 802.11 standard body is currently working on a new amendment, IEEE 802.11 p, to address these concerns. This document is named Wireless Access in Vehicular Environment, also known as WAVE [2]. A Vehicular Ad-Hoc Network, or VANET, is a technology that uses moving cars as nodes in a network to create a mobile network. VANET turns every participating car into a wireless router or node, allowing cars approximately 100–300 m of each other to connect and, in turn, create a network with a wide range. As cars fall out of the signal range and drop out of the network, other cars can join in, connecting

A. Anand (✉) · P. Singh
Chandigarh Engineering College, Landran, Mohali Punjab, India
e-mail: riet_ankita@yahoo.com

P. Singh (✉)
e-mail: singh.parminder08@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_20,

vehicles to one another so that a mobile Internet is created [3] Vehicular Ad hoc network (VANET) is a new form of Mobile Ad hoc Network (MANET) [4].

2 Problem Formulations

As earlier, research is done on Mobile ad hoc network (MANETS)[1] [5]. But with MANETs several handoff problems were difficult to remove. So a new cluster based technique in VANETS is introduced, to remove these handoff problems. And various unpredictable movements that may occur in the network can also be reduced or lessen using this Cluster based technique in VANET.

3 Methodology Used

Scenario is taken as under:

Work is done on Fedora and Windows. In Fedora [6], MOVE is used.
Following is the scenario representing clustered vehicles in VANET environment.
Scenario 1: At the very start, nodes i.e. vehicles are in random order. There is no source node and no destination node.
In this scenario, vehicles are taken as nodes and these vehicles are moving in a cluster. Every vehicle is connected to another vehicle. Firstly a cluster head is decided, which can also be considered as a source node, after that only another nodes can start working. Source node sends data packets to next node and then to next one, resulting data packets to be passed on through various nodes to reach its destination (i.e. the receiving node) (Fig. 1).

This cluster based technique enables nodes or vehicles to remain connected to each other, as even a small problem if any occurs, it may come to know to the whole

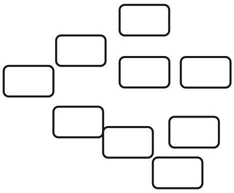


Fig. 1 Initiation of nodes

[1] MANET [5]: A **mobile ad-hoc network** (**MANET**) is a self-configuring infrastructure less network of mobile devices connected by wireless links. *Ad hoc* is Latin and means “for this purpose”. Each device in a MANET is free to move independently in any direction, and will therefore change its links to other devices frequently.

network and the whole network, having several nodes together tries to find out the root cause of the problem, by maintaining good internet connectivity.

Scenario2: After the gathering or generation of all the nodes in the VANET environment, a cluster head is formed, which actually controls the whole process. Cluster head is also considered as a source node (Fig. 2).
Afterwards a simulation for VANET is done.

SIMULATION FOR VANET:
Rapid generation of realistic simulation for VANET.
Here, two models are taken into consideration:

Mobility Model

Traffic Model

Mobility Model: It represents the generation of road map topology and vehicle movement.
In mobility model configuration can be set for roads or the whole scenario with the help of Map configuration editor (Fig. 3).

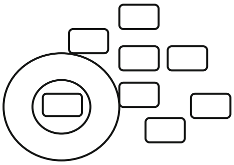
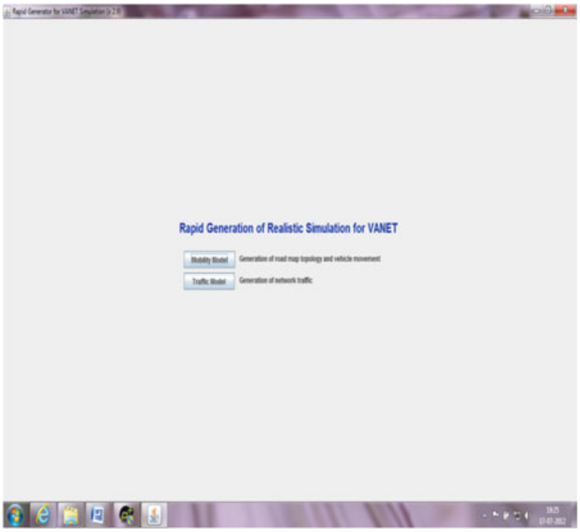


Fig. 2 Formation of a cluster head



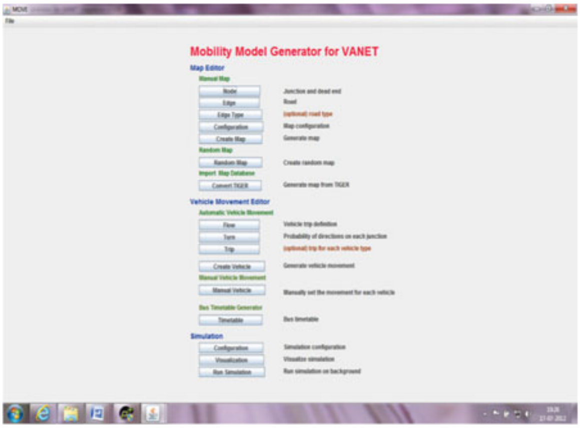
Rapid Generator for VANET Simulation (v 2.0)

Rapid Generation of Realistic Simulation for VANET

Generation of road map topology and vehicle movement

Generation of network traffic

Fig. 3 Rapid generator for VANET



Mobility Model Generator for VANET

Map Editor

Manual Map

Node — Junction and dead end
Edge — Road
Edge Type — (optional) road type
Configuration — Map configuration
Create Map — Generate map

Random Map

Random Map — Create random map

Import Map Database

Convert TIGER — Generate map from TIGER

Vehicle Movement Editor

Automatic Vehicle Movement

Flow — Vehicle trip definition
Turn — Probability of directions on each junction
Trip — (optional) trip for each vehicle type

Create Vehicle — Generate vehicle movement

Manual Vehicle Movement

Manual Vehicle — Manually set the movement for each vehicle

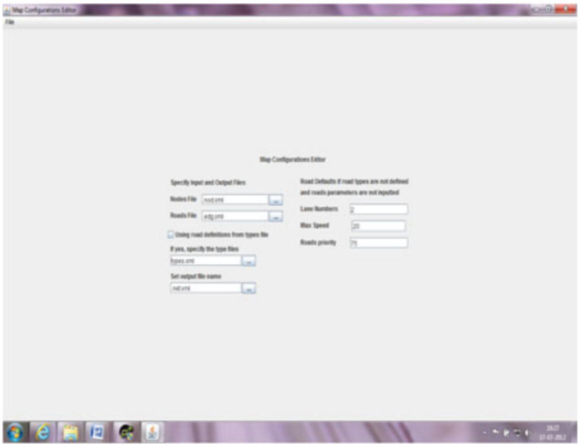
Bus Timetable Generator

Timetable — Bus timetable

Simulation

Configuration — Simulation configuration
Visualization — Visualize simulation
Run Simulation — Run simulation on background

Fig. 4 Mobility model



Map Configurations Editor

Specify Input and Output Files

Nodes File: nod.xml
Roads File: edg.xml

□ Using road definitions from types file

If yes, specify the type files: types.xml

Set output file name: net.xml

Road Defaults if road types are not defined and roads parameters are not inputted

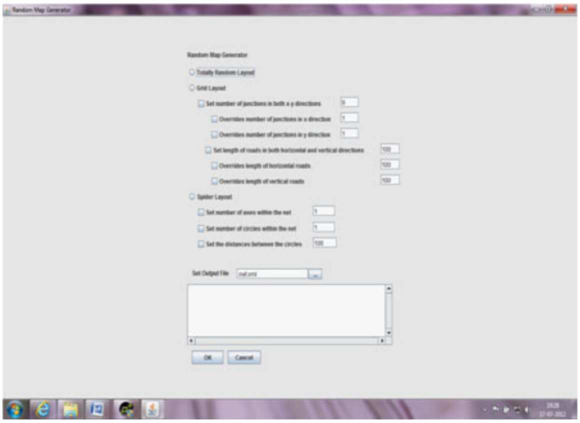
Lane Numbers: 2
Max Speed: 20
Roads priority: 75

Fig. 5 Map configuration Editor

Map configuration Editor: It specifies the input and the output files and the road defaults if road types are not defined and road parameters if they are not inputted (Fig. 4).
Random map generator: It specifies the total random layout which includes Grid layout and the spider layout (Fig. 5).
Traffic Model: It basically represents the Generation of network traffic.
Traffic model shows two mobility for Ns2 i.e.

Static Mobility

Dynamic Mobility



Random Map Generator

○ Totally Random Layout

○ Grid Layout

□ Set number of junctions in both x y directions 9

□ Overrides number of junctions in x direction 1

□ Overrides number of junctions in y direction 1

□ Set length of roads in both horizontal and vertical directions 100

□ Overrides length of horizontal roads 100

□ Overrides length of vertical roads 100

○ Spider Layout

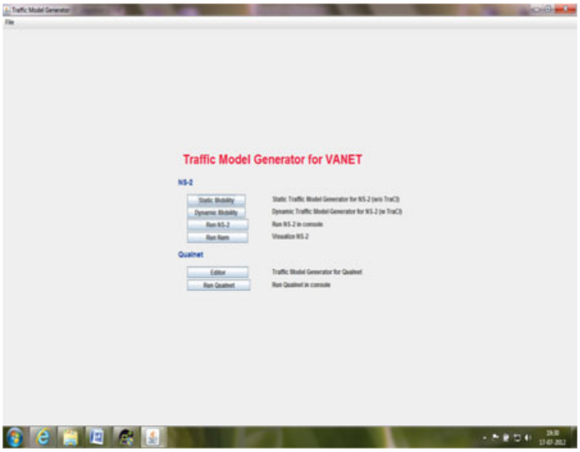
□ Set number of axes within the net 1

□ Set number of circles within the net 1

□ Set the distances between the circles 100

Set Output File net.xml

Fig. 6 Random Map Generator



Traffic Model Generator for VANET

NS-2

Static Traffic Model Generator for NS-2 (w/o TraCI)

Dynamic Traffic Model Generator for NS-2 (w TraCI)

Run NS-2 in console

Visualize NS-2

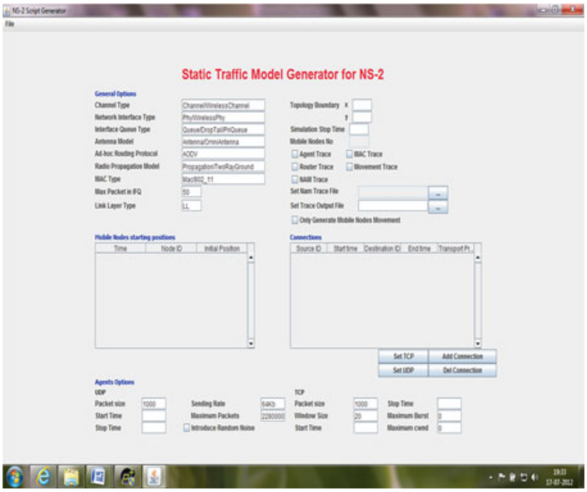
Qualnet

Traffic Model Generator for Qualnet

Run Qualnet in console

Fig. 7 Traffic model

Static Mobility: It is a static traffic model generator for Ns2. It includes general options like channel type, network interface type, interface queue type, antenna model, ad hoc routing protocol, radio propagating model, Mac type, Max packet in, link layer type and mobile node starting positions, agent options, and connections (Fig. 6).
Dynamic Mobility: It is a dynamic traffic model generator for Ns2. It also includes general options like channel type, network interface type, interface queue type, antenna model, ad hoc routing protocol, radio propagating model, Mac type, Max packet, link layer type and mobile node starting positions, agent options, and connections for dynamic mobility of vehicles (Fig. 7).



Static Traffic Model Generator for NS-2

General Options
Channel Type: Channel/WirelessChannel
Network Interface Type: Phy/WirelessPhy
Interface Queue Type: Queue/DropTail/PriQueue
Antenna Model: Antenna/OmniAntenna
Ad-hoc Routing Protocol: AODV
Radio Propagation Model: Propagation/TwoRayGround
MAC Type: Mac/802_11
Max Packet in IFQ: 50
Link Layer Type: LL

Topology Boundary x:
y:
Simulation Stop Time:
Mobile Nodes No:
□ Agent Trace □ MAC Trace
□ Router Trace □ Movement Trace
□ NAM Trace
Set Nam Trace File:
Set Trace Output File:
□ Only Generate Mobile Nodes Movement

Mobile Nodes starting positions

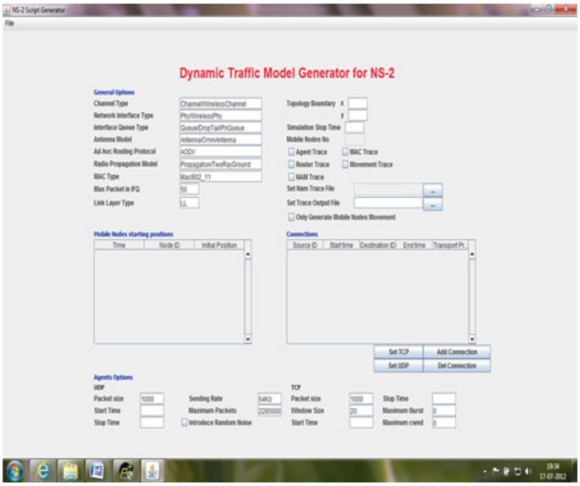
Time	Node ID	Initial Position

Connections

Source ID	Start time	Destination ID	End time	Transport Pr

Agents Options
UDP
Packet size: 1000
Start Time:
Stop Time:
Sending Rate: 64Kb
Maximum Packets: 2280000
□ Introduce Random Noise
TCP
Packet size: 1000
Window Size: 20
Start Time:
Stop Time:
Maximum Burst: 0
Maximum cwnd: 0

Fig. 8 Static model generator for VANET



Dynamic Traffic Model Generator for NS-2

General Options
Channel Type: Channel/WirelessChannel
Network Interface Type: Phy/WirelessPhy
Interface Queue Type: Queue/DropTail/PriQueue
Antenna Model: Antenna/OmniAntenna
Ad-hoc Routing Protocol: AODV
Radio Propagation Model: Propagation/TwoRayGround
MAC Type: Mac/802_11
Max Packet in IFQ: 50
Link Layer Type: LL

Topology Boundary x:
y:
Simulation Stop Time:
Mobile Nodes No:
□ Agent Trace □ MAC Trace
□ Router Trace □ Movement Trace
□ NAM Trace
Set Nam Trace File:
Set Trace Output File:
□ Only Generate Mobile Nodes Movement

Mobile Nodes starting positions

Time	Node ID	Initial Position

Connections

Source ID	Start time	Destination ID	End time	Transport Pr

Agents Options
UDP
Packet size: 1000
Start Time:
Stop Time:
Sending Rate: 64Kb
Maximum Packets: 2280000
□ Introduce Random Noise
TCP
Packet size: 1000
Window Size: 20
Start Time:
Stop Time:
Maximum Burst: 0
Maximum cwnd: 0

Fig. 9 Dynamic model generator for VANET

4 Results and Discussion

Results come out in the form of Generated packets and dropping packets, for Clustered vehicles in VANET environment (Fig. 8).
(i) Generated packets: It tells us about how many packets are generated as well as their packet identification. Here is a graph of generated packets is plotted on the XY axis. When a packet is transmitted, some packet delay comes. Like if 1000 packets

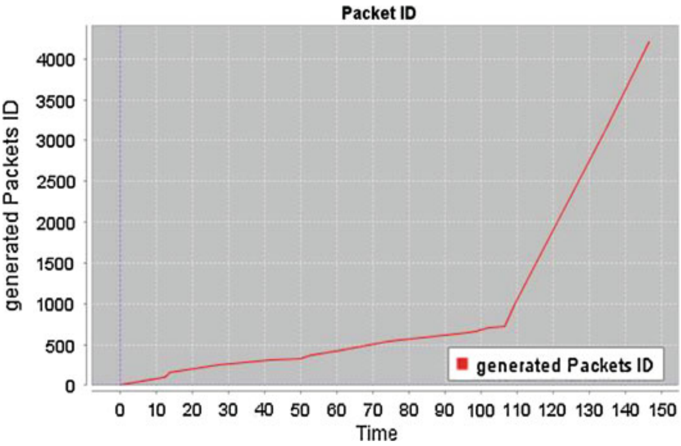


Fig. 10 Generated packets

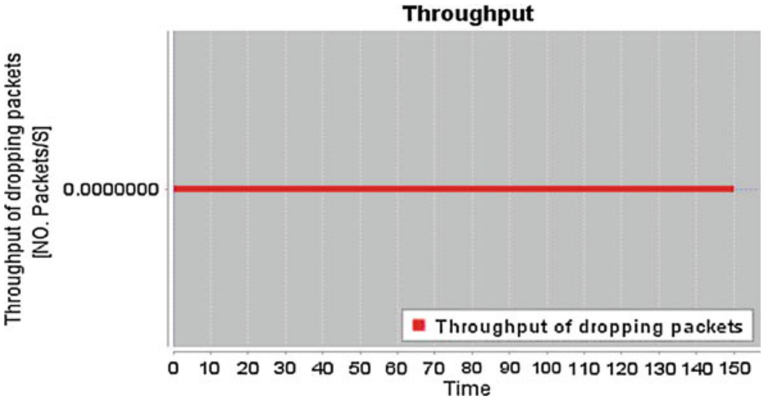


Fig. 11 Throughput of dropping packets

are to be transmitted, all the 1000 can not be transmitted at once, but the packets can be transmitted in instalments, thereby resulting in the packet delay (Fig. 9).
(ii) Throughput of dropping packets: It tells the throughput of the dropped packets on XY axis, with throughput of dropping packets on X axis and throughput of dropping packets (No. Of packets) on Y axis (Fig. 10).

5 Conclusion

In this paper, analysis has been done on VANET, using the Cluster based technique in terms of Generated packets in the network and the throughput of the dropping packets that have been dropped out of the network. And it is concluded that VANET with this

Cluster based technique proves to be helpful in removing the handoff problems that usually occurred in MANETS and were not possible to remove. This clustered based technique in VANETs also helps a great deal in removing or lessen the unpredictable movements in the network, which sometimes may cause dangerous problems like dropping down the whole network for any small mistake. Each vehicle in the VANET with cluster based technique is connected to one another, resulting in a good internet connectivity (Fig. 11).

Acknowledgments Foremost of all, I express my sincere indebtedness to the "Almighty", for bestowing me with all the favourable circumstances and kept me in high spirits. I am very grateful to my advisor, Er Parminder Singh, Asst Prof. at Chandigarh Engineering College, Landran for his patience, encouragement, and many fruitful discussions from the very early stage of this research.. Finally, I wish to express my appreciation to my family for their continuous love and encouragement, for always believing in me, for never failing to provide all the support, and for coping with the pressure that naturally comes with such endeavour.

References

1. "IEEE802.11p", http://en.wikipedia.org/wiki/IEEE_802.11p
2. Jiang D, Delgrossi L (2008) Mercedes-Benz Research & Development North America, Inc., "IEEE 802.11p: Towards an International Standard for Wireless Access in Vehicular Environments"
3. "VANET", http://en.wikipedia.org/wiki/Vehicular_ad-hoc_network
4. Lo N-W, Tsai H-C (2007) "Illusion Attack on VANET Applications—A Message Plausibility Problem"
5. "Manet", http://en.wikipedia.org/wiki/Mobile_ad_hoc_network
6. "Fedora", http://fedora-commons.org/

Performance Evaluation of TCP Congestion Control Variants Using Ad Hoc On-Demand Distance Vector Routing

Mayank Kumar Goyal, Punit Gupta and Vinita Chauhan

Abstract A mobile ad hoc network is a collection of wireless mobile nodes dynamically forming a temporary network without the use of any existing network infrastructure. This paper presents the comparative performance analysis of Ad hoc On Demand Distance Vector routing for mobile ad hoc networks with different versions of TCP congestion control. A network simulator has been used for performance evaluation of Ad Hoc On-Demand Distance Vector Routing in this paper. To compare the performance of this protocol, the simulation results were analyzed using Quality of Service metrics such as throughput, packet delivery ratio, packet loss and average end to-end delay. The results presented in this paper clearly indicate that the Ad hoc On Demand Distance Vector routing achieves maximum throughput, higher packet delivery ratio and less average end to end delay when TCP congestion control agent used is TCP Vegas.

Keywords AODV · NS2 · TCP · Tahoe · Reno · New Reno · Vegas · MANET

M. K. Goyal(✉)
Department of CS/IT Engineering, Sharda University, Greater Noida, India
e-mail: mayankrkgit@gmail.com

P. Gupta
Department of Computer Science Engineering, ABES Engineering College, Ghaziabad, India
e-mail: punitg07@gmail.com

V. Chauhan
Department of Computer Science Engineering, IMS Engineering College, Ghaziabad, India
e-mail: vini.20m@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_21,

1 Introduction

A mobile ad hoc network is a kind of wireless ad-hoc network, and is a self-configuring network of mobile routers connected by wireless links—the union of which form an arbitrary topology [1, 11]. The routers are free to move randomly and organize themselves arbitrarily thus, the network's wireless topology may change rapidly and unpredictably [10]. Routing is the process of moving a data packet from source to destination. Routing is usually performed by a dedicated device called a router [5, 6]. TCP is the transport layer protocol used to provide a reliable end-to-end data transmission.

This paper presents the comparative performance analysis of Ad hoc On Demand Distance Vector routing for mobile ad hoc networks with most used versions of TCP congestion control i.e. TCP Tahoe, TCP Reno, TCP New Reno and TCP Vegas. For this purpose, the simulation results were analyzed using Quality of Service metrics such as throughput, packet delivery ratio, packet loss and average end-to-end delay.

This paper is organized as follows. Section 2 describes Ad Hoc On-Demand Distance Vector Routing (AODV), Sect. 3 gives TCP Congestion control mechanism. In Sect. 4, Performance metrics & simulation environment are presented. Section 5 describes the Results.

2 Ad Hoc On-Demand Distance Vector Routing (AODV)

The Ad hoc On Demand Distance Vector routing algorithm is a routing protocol designed for ad hoc mobile networks [1, 3]. It is capable of both unicast and multicast routing [12]. It builds routes between nodes only as desired by source nodes [2, 4, 7]. It makes routes using a route request/route reply query cycle. When a source node wants a route to a destination for which it does not already have a route, it broadcasts a route request (RREQ) packet across the network. Nodes receiving this data packet update their information for the source node and set up backwards pointers to the source node in the route tables. In addition to the source node's IP address, current sequence number, and broadcast ID, the RREQ also contains the most recent sequence number for the destination of which the source node is aware of. A node receiving the RREQ may send a route reply (RREP) if it is either the destination or if it has number more than or equal to that contained in the RREQ. If this is the case, it unicasts a RREP back to the source.

3 TCP Congestion Control Mechanism

3.1 Tahoe

In TCP Tahoe, the sender maintains a congestion window (CWD) that limits the no. of packets that it is allowed to send into the network without waiting for acknowledgment. The congestion control is a two-phase control mechanism: a slow-start phase and a congestion avoidance phase. The TCP sender dynamically increases/decreases the congestion window size according to the congestion level, which is conjectured by packet losses. Switching from the first phase to second phase depends on the slow-start threshold (ssthres). Slow starts suggest that the sender set the congestion window to 1 and then for each ACK received it increase the CWD by 1. So in the first round trip time (RTT) it sends one packet, in the second it sends two and in the third it sends four [13, 14].

3.2 RENO

TCP Reno adds some intelligence to TCP Tahoe so that the packets which are lost are detected earlier and the pipeline is not vacant every time a packet is lost. Reno requires that receiving immediate acknowledgement whenever a segment is received. The logic behind this is that whenever it receive a duplicate acknowledgment, then his duplicate acknowledgment could have been received if the next segment in sequence expected, has been delayed in the network and the segments reached there out of order or else that the packet is lost. If it receives a number of duplicate acknowledgements then that means that sufficient time have passed and even if the segment had taken a longer path, it should have gotten to the receiver by now. There is a very high probability that it was lost. So Reno suggests an algorithm called ‘Fast Re-Transmit’ [9].

3.3 NEW-RENO

TCP New RENO is able to detect multiple packet losses and thus is much more effective than RENO in the case of multiple packet losses. Like Reno, New-Reno also enters into fast-retransmit when it receives multiple duplicate packets, it doesn’t exit fast-recovery until all the data which was out standing at the time it entered fast recovery is acknowledged. Thus it overcomes the problem faced by Reno of reducing the CWD multiples times. The fast-transmit phase is the same as in Reno. The difference in the fast recovery phase which allows for multiple re-transmissions in new-Reno [8, 11].

3.4 VEGAS

TCP Vegas detects congestion at an incipient stage based on increasing Round-Trip Time (RTT) values of the packets in the connection unlike other flavors like Reno, New Reno, etc., which detect congestion only after it has actually happened via packet drops. The working depends heavily on accurate calculation of the Base RTT value [10]. If it is too small then throughput of the connection will be less than the bandwidth available while if the value is too large then it will overrun the connection. One interesting fact is that when Vegas is inter-operated with other versions like Reno, performance of Vegas degrades because Vegas reduces its sending rate before Reno as it detects congestion early and hence gives greater bandwidth to co-existing TCP Reno flows [4].

4 Performance Metrics

End-to-End Delay: The average time interval consumed between the generation of a packet in a source node and the successfully delivery of the packet at the destination node.

Packet Delivery Fraction: The ratio of the number of data packets successfully delivered to all destination nodes and the number of data packets generated by all source nodes.

Number of Packets dropped: The number of data packets that are not successfully delivered to the destination.

Throughput: Throughput is the number of packet arriving at the sink per millisecond.

5 Results

Simulation results are taken for 25 nodes and are shown in Table 1. Table 1 represents the values for performance metrics considered above for Ad hoc On Demand Distance Vector routing when implemented with different versions of TCP congestion control mechanism.

5.1 Packet Delivery Ratio Comparison

The packet delivery ratio for Ad hoc On Demand Distance Vector routing achieved similar values when implemented with TCP Tahoe, TCP Reno and TCP New Reno. AODV achieved maximum packet delivery ratio when implemented with TCP Vegas.

5.2 Throughput Comparison

The throughput for Ad hoc On Demand Distance Vector routing achieved similar values when implemented with TCP Tahoe, TCP Reno and TCP NewReno. AODV achieved maximum throughput when implemented with TCP Vegas.

5.3 Avg. End-to-End Delay Comparison

Avg. End-to-End Delay for Ad hoc On Demand Distance Vector routing achieved similar values when implemented with TCP Tahoe, TCP Reno and TCP NewReno. AODV achieved minimum Avg. End-to-End Delay when implemented with TCP Vegas.

5.4 Simulation Graph for AODV TCP Congestion Control Mechanism

Simulation graph for performance analysis of Ad Hoc On-Demand Distance Vector Routing with different versions of TCP Congestion Control has been plotted using Xgraph utility of ns2 and Figs. 1, 2, 3 and 4 represents it below.

Figures 1 and 2 represents identical results for AODV implementation with TCP Tahoe & TCP Reno.

Figures 3 and 4 does not represent identical results for AODV implementation. Table 1 clearly indicates that AODV implementation with TCP Vegas achieved maximum throughput, higher packet delivery ratio and less average end to end delay.

Table 1 Simulation Results for 25 Nodes

Performance metric	Tahoe	Reno	New Reno	Vegas
Start time (s)	10	10	10	10
Stop time (s)	60	60	60	60
Generated packets	1258	1258	1258	1289
Received packets	623	623	623	641
Packet delivery ratio	49.5231	49.5231	49.5231	49.728
Total dropped packets	12	12	12	6
Avg. End-to-End delay (ms)	121.274	121.274	121.274	71.033
Avg. Throughput(kbps)	237.41	237.41	237.41	255.11

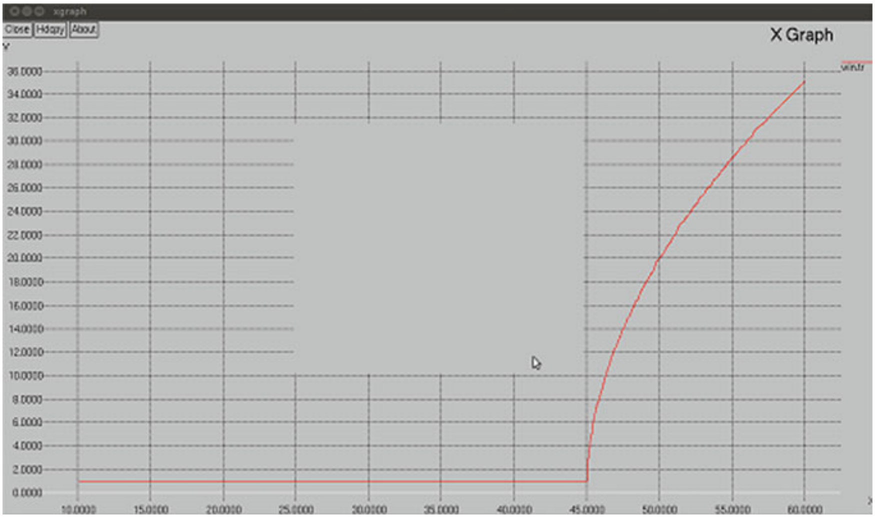


Fig. 1 Xgraph for AODV implementation with TCP Tahoe

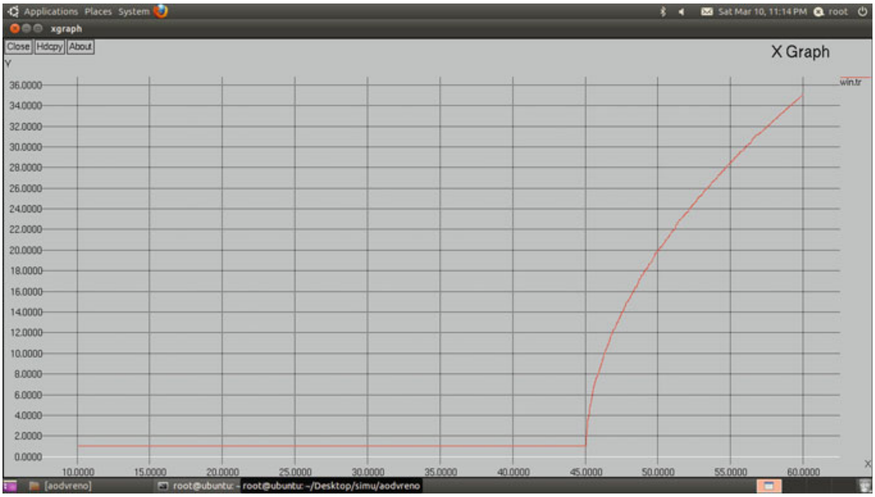


Fig. 2 Xgraph for AODV implementation with TCP Reno

6 Conclusion

The average bandwidth available to different versions of TCP congestion control mechanism are not the same. But for this network simulation environment, we prefer to adopt TCP Vegas because the average bandwidth is less than other TCP versions. TCP-Vegas experienced an inaccuracy problem of Base RTT due to frequent path

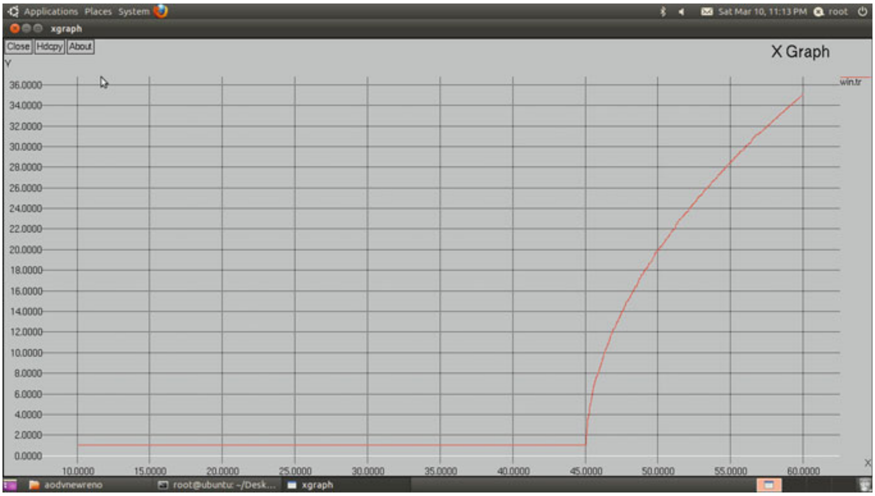


Fig. 3 Xgraph for AODV implementation with TCP New Reno

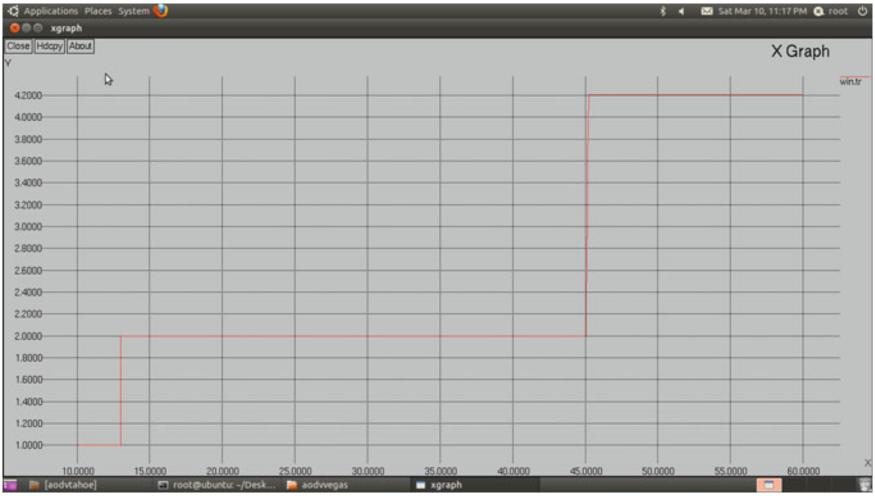


Fig. 4 Xgraph for AODV implementation with TCP Vegas

change caused by node mobility causing performance degradation. Therefore, some modification in order to estimate an exact Base RTT over a new path is required to improve the TCP performance for MANET. On other side, the aggressive window increasing attitude of TCP-Reno also makes throughput decrease caused by hidden node problem.Ad hoc On Demand Distance Vector routing achieved its best performance metrics i.e. maximum throughput, higher packet delivery ratio and less average end to end delay when TCP congestion control agent used was TCP Vegas.

References

1. Abdullah A, Ramly N, Muhammed A, Derahman MN (2008) Performance comparison study of routing protocols for mobile grid environment. IJCSNS Int J Comput Sci Netw Sec 8(2):82–88
2. Alexander Z (2003) Performance Evaluation of AODV Routing Protocol: Real-Life Measurements. SCC
3. Altman E, Jimenez T (2003) NS Simulator for Beginners. Lecture notes. Univ.de Los Andes, Merida, Venezuela and ESSI. Sophia-Antipolis, France
4. Chakeres ID, Belding-Royer EM (2004) AODV Routing Protocol Implementation Design. In: Proceedings of DCS, pp 698–703
5. Johnson D (2003) Dynamic source routing for mobile Ad Hoc Networks. IEFT MANET Draft
6. Johnson DB, Maltz DA, Broch J (2001) DSR: The dynamic source routing protocol for multi-Hop wireless Ad Hoc networks. Internet-Draft, draft-ietfmanet-dsr-00. Txt, 1978
7. Johnson DB, Maltz DA (1996) “Dynamic source routing in adhoc wireless networks”. In: lmielinski T, Kmh H (eds) Mobile computing, Kluwer Academic, ch. 5, 1996
8. Johnson D, Maltz D, Broch J (2001) DSR. In: Perkins C (ed) The dynamic source routing protocol for multihop wireless Ad Hoc networks in Ad Hoc networking . Chapter 5:139–172
9. Lee H, Lee S, Choi Y (2001) “The influence of the large bandwidth-delay product on TCP Reno, NewReno, and SACK”. In: Proceedings Information Networking Conference, Oita, Japan, 327–334, Feb 2001
10. Murthy CSR, Manoj BS (2004) Ad Hoc Wireless Networks: Architecture and Protocols, ch. Routing Protocols for Ad Hoc Wireless Networks, Prentice Hall Communications Engineering and Emerging Technologies Series, New Jersey: PrenticeHall Professional Technical Reference, pp. 299–364, 2004
11. Nesargi S, Prakash R (2002) “MANETconf: Configuration of Hosts in a Mobile Ad Hoc Networks”, 21st Annual Joint Conf. of the IEEE Computer and, Communications Societies (INFOCOM)
12. Perkins CE, Royer EM (1999) “Ad Hoc on-demand distance vector routing”. In: Proceedings of the 2nd IEEE Workshop on Mobile Computing Syslems and Applications, 90–100 Feb 1999
13. Staub T (2004) Ad-hoc and hybrid networks: performance comparison of MANET routing protocols in Ad-hoc and hybrid networks. University of Berne, Switzerland, Institute of computer science and applied mathematics, pp 1–38
14. Zeng WG, Trajkovic Lj (2005) TCP Packet control for wireless networks. In: Proceedings IEEE Int. Conf. on Wireless and Mobile Computing. Networking and Communications 2:196–203

Performance Analysis of Dynamic Source Routing for Ad-Hoc Networks Using Active Packet

Manish Bhardwaj, Naresh Sharma and Ruchika Saini

Abstract DSR (Dynamic Source Routing) is an on-demand routing method for ad-hoc wireless networks that floods route requests when the route is needed. Route caches in DSR are used to reduce flooding of route requests. But with the increase in network size and node mobility, cached routes quickly become stale or inefficient. The consequence is a huge number of routing packets in the network that consume significant bandwidth and computing resources. This paper presents a deployable active network approach to this route cache problem. In our method, an active packet roams around the network, and collects network topology information. By checking the content of this active packet when it passes through, network nodes are able to positively update their route caches. Thus cache miss rates decrease and the route discovery flooding is reduced. Simulation results show that our method reduces the miss rates by up to 65 % and routing packet numbers by up to 49 %.

Keywords Dsr · Manet · Ns 2

1 Introduction

DSR [1] is a widely used on-demand ad-hoc network routing strategy that uses route caches. DSR floods route requests to find a route that is needed. Since the cost of flooding is high, DSR maintains route caches to store routes that have been found via flooding or through promiscuous overhearing. DSR route caches have two

M. Bhardwaj (✉) · N. Sharma · R. Saini
Department of Computer Science and Engineering, SRM University, Modinagar, India
e-mail: aapkaapna13@gmail.com

N. Sharma
e-mail: nrssharma@gmail.com

R. Saini
e-mail: ruchika.sre@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_22,

disadvantages. First, DSR takes no measures to avoid flooding for a route to a "new" destination, which has not been overheard yet. DSR has to flood route requests to get the route and save it in route caches. Second, entries in route caches quickly become invalid or inefficient when node mobility is high. When an inefficient route is used, data packets suffer unnecessary delays; when an invalid route is followed, route failures will trigger flooding, creating additional latency for data packets.

Although table-driven routing algorithms [2–9] could overcome the new-destination problem, they entail too much overhead. In these methods, each node needs to either exchange its view of the whole network topology with its neighbors or flood its link states to the whole network. The involved routing overhead is significant.

There are proposals for tackling the stale-cache problem [10–12]. In [10] a link-based cache structure is proposed that encodes more topology information than a traditional path-based cache structure. In [11] the route failure notification mechanism is extended so that all nodes having the related broken links are informed to remove them. The paper also introduces negative caches to avoid using broken paths. In [12], authors investigate ways to validate and shorten cached routes. These proposed schemes are relatively complex. For example, in [10], maintaining the link-based cache is quite difficult in a high mobility network. In [12], it is unclear how to make decisions such as what subset of paths should be validated, and when the validation should happen.

In this paper, we propose a solution to both of the above problems of route caches in DSR using active network techniques [13, 14]. In our approach, an active helper is dynamically loaded upon arrival of an active packet, which is originated from an arbitrarily-chosen node. Driven by the active helper, the active packet travels through the network, collects topology information, and has nodes update their route caches. Route entries then contain new routes reflecting the most recent topology changes. Thus both problems are taken care of and flooding is reduced. This paper shows through simulations that improvements from this approach are significant. Besides, the active helper is deliberately designed to be an independent module and does not change any existing aspect of the DSR module. Thus, its deployment in current ad-hoc networks should be relatively easy.

This paper is organized as follows. In Sect. 2, two measurements are given to quantify DSR performance. Then the active approach is described in Sect. 3 to improve DSR. Sect. 4 presents simulation results. Finally a conclusion is given in Sect. 5.

2 Route Miss Rate and Routing Packet Number

We define that a route cache miss happens not only when the needed route is not present in caches, but also when a previous hit has turned out to be wrong. Both cases trigger the route request flooding. Thus the *miss rate* quantity gives the degree of the flooding. The *routing packet number* counts all routing-related packets injected by all nodes in an ad-hoc network in order to send a certain number of data packets. This quantity reflects the overhead of the routing protocol.

We use ns-2 [15] simulations to measure these quantities. We simulated different sizes and mobility of ad-hoc networks to see how miss rates and routing packet numbers change in different scenarios. In each network, the moving patterns of nodes follow the Random Waypoint Motion Model [16]. In this model, each node chooses a destination with a uniformly random distribution over the network area, moves there at a speed uniformly between 0 and the maximum allowable velocity, stays there for a specified pause time, and then repeats the behavior. During each simulated session, nodes are uniformly randomly chosen to be the senders or receivers, and carry out CBR communications for a uniformly randomly chosen time. The number of CBR connections in each session is proportional to the number of nodes.

We simulated networks with 3 different sizes and 5 different pause times. The first size is 20 nodes within a 670 × 670 m area, the second is 50 nodes within a 1500 × 700 m area, and the third is 100 nodes within 2000 × 1000 m area. The tested pause times are 300, 100, 50, 20 and 0 ms. For the 20 node network, the maximum connection number that could be built during the simulation session is 40, that number is 100 for the 50 node network, and 200 for the 100 node network. The maximum moving speed is 20 m/s. The CBR rate is 5 packets per second. The simulation time is 500 s. The results are shown in Fig. 1.

From Fig. 1, in each network, the miss rate generally increases with increasing in mobility (pause times changed from 300 to 0 ms). DSR performs well in small-size networks. However, when the network size gets large (about 100 nodes within 2000 × 1000 m area), the average route path becomes longer (the nominal transmission range is 250 m). Since a longer path is more likely to get invalid than a shorter one, the miss rates go up significantly, and so does flooding. As a result, the routing packet numbers also increases.

Note that the increase in routing packet numbers is more apparent than the increase in miss rates. For example, the miss rate goes up from 14 to 21 % when the network size changes from 50 to 100 with a fixed pause time of 20 ms, but the routing packet number (per data packet) increases dramatically from 9 to 30. The reason is flooding

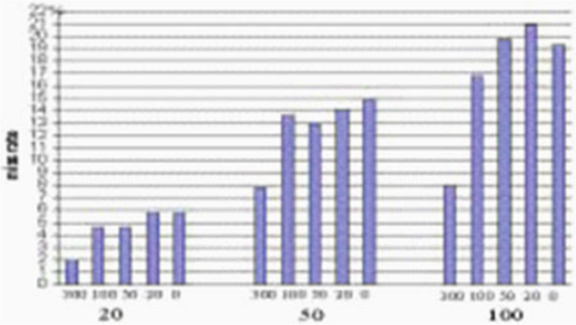


Fig. 1 Route miss rates (percent)

goes much further in a large network than in a small network. Thus the same miss rate entails much higher overhead in a larger network.

3 Active Network Approach

Our approach to reducing miss rates and routing packet numbers is to keep route caches updated in an active way. Active networks [13, 14], because of their programmability and dynamic-module-loading property, provide a good solution. There is an active helper module that operates on a predefined active packet. This module is independent of the existing DSR module. It is loaded only when the active packet arrives or originates at a node. Then this helper works in the background without interrupting normal operations of the DSR module. This independent design makes it easy to deploy our mechanism.

An active packet is periodically generated by a randomly-chosen node. The basic idea is to drive the active packet to visit each node twice. The first visit is to obtain topology information of the network; the second visit is to update route caches against the obtained information. There is a marker field in the active packet header to indicate if the packet is in the first or the second visit. The payload of the active packet is a connection matrix for the network topology. Upon being dynamically loaded, the active helper checks the header of the active packet. If the packet is on the first visit, the helper checks the nodes neighbor information and updates the connection matrix of the packet before sending it out. If the active packet is in its second visit, the helper makes use of its payload to validate and update the nodes route cache positively. The active helper is also responsible for changing the marker field when the packet finishes its first visit and freeing the packet when the second visit is completed too. Periodically, the active helper and the active packet are triggered and route caches keep being updated. The following sections describe the active helper with more details.

3.1 Visiting Nodes

It is assumed that each node knows its neighbors. This assumption is reasonable because a mobile node can issue beacons to detect neighbors within a nominal time. The algorithm also assumes that routes are bi-directional, which is the case for the typical wireless MAC protocol 802.11 [17].

Upon being created, an active packet begins to visit nodes in a depth-first order. During its first visit, the connection matrix of the active packet keeps being added to the topology information until the first visit is completed. Due to the depth-first property of the visit, the next node to be visited is generally a neighbor of the node that is currently being visited, except when backtracking happens. In this case, the next node is a neighbor of some node previously visited. A mechanism is needed to have the active packet go back to the previous node so that the visit can continue. Our method is to let the active helper compute a route on the fly from the current node to the previous one according to the information in the connection matrix. This way, less state information is kept for backtracking.

One question is what will happen if the network topology changes in the process of visiting. There are two possibilities; one is links between nodes the active packet has visited still exist, but there are also some new links appearing (including the situation when new nodes come into the network). In this case, the connection matrix, although not a complete reflection of links, is still valid. Thus the paths between visited nodes will still be valid, although may not be optimal. The other possibility is that some links are lost. In this case, the information about these links in the connection matrix is incorrect and paths containing these links are not valid. But the information in the connection matrix about other links should still be valid and useful. To address the above problems, we take measures mentioned in Sect. 3. D to make the visiting of nodes fast compared to node movements.

If there are partitions preexisting in the network before the active packet visit happens, nodes in the same partition as the initiating node get visited, while nodes in other partitions are generally not visited. This is desirable since those visited nodes still get benefits while others remain untouched and don't load the active helper at all.

3.2 Updating Route Caches

During the second visit of the active packet, the active helper validates and updates the route cache of a node according to the connection matrix in the packet. In the validation phase, each routing entry is checked against the connection matrix and those that disagree are removed. This way, the flooding coming from route failures is reduced. In the updating phase, the active helper adds routes learned from the active packet. There are two ways to decide which routes should be added into the cache. The first approach is *conservative*; the node only adds those routes with destinations identical to the ones that old route entries had before the cache was checked. The idea is to update only preexisting routes. This strategy consumes little space in the route cache. The second approach is *proactive*; besides updating preexisting routes, nodes also add new routes that are valid. This strategy needs more cache space, but provides routes for future use, thus reducing new route request flooding. We implemented both updating methods.

3.3 Timers

There are two timers used in our approach. The initial timer is used to initiate the visits periodically; each time the initial timer expires, an active packet is generated and is sent out. The other is called maintenance timer. When the maintenance timer expires, route entries added by previous active packets are cleared. This timer has an interval that is a little smaller than the initial timer. It is useful when network partitions happen between visits. The nodes in the disconnected partitions automatically clear the entries left by old active packets, avoiding negative impacts of these possible stale routes.

There is a tradeoff in choosing the interval of the initial timer. The shorter the interval, the more frequent the updates, and the more recent the route cache entries. Thus the route flooding should be less frequent. However, a shorter interval also means that there are more active packets roaming around the network that consume network bandwidth. The value of the initial timer should be chosen to keep both miss rates and routing overhead low.

3.4 Visiting Efficiency

We took some measures to improve the visiting efficiency in our implementations. First, there is a function called TAP in DSR to allow nodes to promiscuously hear passed packets for the purpose of finding useful routes. In our implementation, we don't allow TAP on the active packets. Thus the TAP processing is saved for active packets and they move more quickly up along the network stack space. Furthermore, we put all out-going active packets in the head of the priority queue on the network interface. The priority order in the queue from high to low becomes active packets, DSR packets and then data packets. The active packets get the highest priority which saves their waiting time.

There are other possible measures. For example, the active packet can be restricted to visit each node only once, instead of twice. The active helper then updates route caches along the way of the visit by using partial topology information in the connection matrix. Another possibility is to make active packets location-aware and limit how far they can go. We are in the process of studying these measures.

4 Simulation Results

To compare our protocol with the classical DSR routing, we call DSR with our active helper ACT. The same set of simulations as described in Sect. 2 was run under both DSR and ACT, for the conservative and proactive ACT. First we tested how much gain there is for the conservative method to update route caches. Figure 2 shows results of running DSR and ACT on three scenarios, representing different network sizes and pause times.

From Fig. 2, conservative ACT indeed improves upon DSR. This is because updated route caches reflect more recent information for preexisting routes. The flooding triggered by stale routes is thus relieved. The miss rates get a slight improvement. As mentioned before, a slight decrease of miss rates can lead to significant

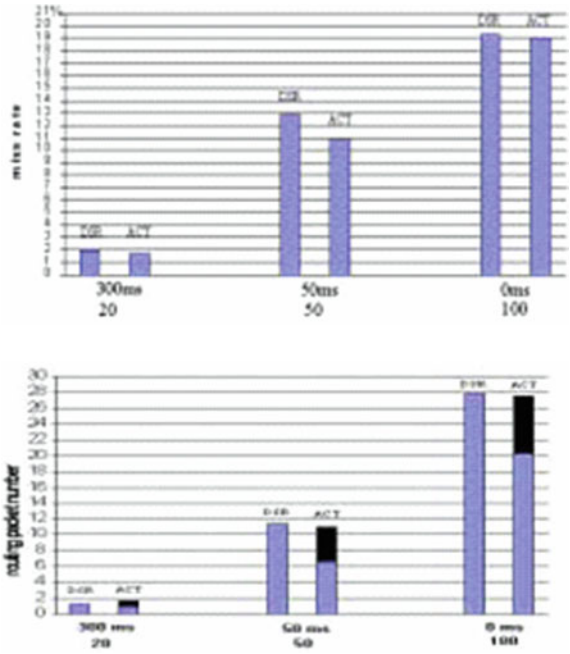


Fig. 2 Miss rates and routing packet numbers (per data packet) from the conservative ACT

savings for the flooding in a large network. The DSR packets get a good portion of savings (shown as the gray parts in the right half of the figure). But because the ACT packets should also be counted as routing packets (shown as the dark parts), the total savings of routing packets are not high.

On the other hand, if route caches are updated with new routes that may be used in future, the flooding triggered by new requests will also be reduced. For this purpose, we ran the proactive ACT for 3 different size networks, each with 5 different pause times. The results are shown in Fig. 3 through Fig. 4.

There are several points noticeable from the above figures. First, the proactive ACT performs much better than the conservative one. For example, for the 100 node network with the pause time of 0 ms, the conservative ACT improves the miss rate only by 1.7 % and the routing packet number only by 1 % (Fig. 2), while these two values for the proactive ACT are 33 and 47 % respectively (Fig. 4). This is because the larger portion of DSR flooding is saved by also caching future routes. For example, in the scenario of the 100-node network with pause time 100 ms, the proactive ACT drops the DSR packet number from 28.2 to 8.1, saving more than 71 % DSR routing packets. Although the active packets add to the number of total routing packets, the saving is still a lot.

For small and medium networks, the reduced miss rates are more impressive, while for larger networks, the savings for routing overhead are more impressive. The reason is that it is more difficult to have the cache entries remain valid with the increase in network size. One link change in a large network leads to a lot more

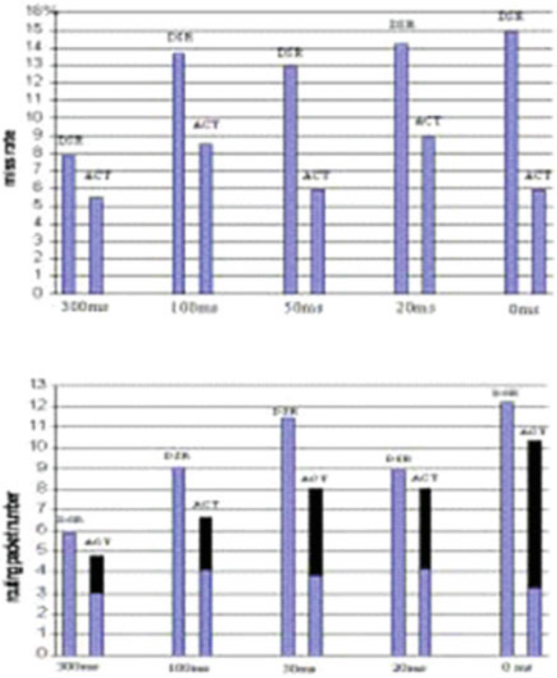


Fig. 3 Miss rates and route packet numbers (per data packet) from the proactive ACT for the 50-node network

route changes. The same value of a reduction of the flooding rate has more apparent effects in a large network than a small network. In simulations, we get least amount of improvement in routing packet number for the 20-node network. For many scenarios of this network, the packet number even goes up slightly although the miss rate improvements are good.

The largest improvement for the miss rate is in the scenario of 50-node network with pause time 0 ms, which is 65 %. The largest improvement for the routing packet number comes from the scenario of 100-node network with pause time 0 ms, which is 49 %.

5 Conclusions

In this paper, we presented an active network approach for improving route failures and overhead with DSR. This method uses an active packet that periodically visits all nodes it can reach to get network topology information and then uses this information to validate and update the cached routes. With active networking, we not only adjust existing routes, but also cache future routes based on the topology information. Thus both route request flooding for the stale routes and new routes are reduced. The reduction in the flooding rate also significantly reduces the routing overhead.

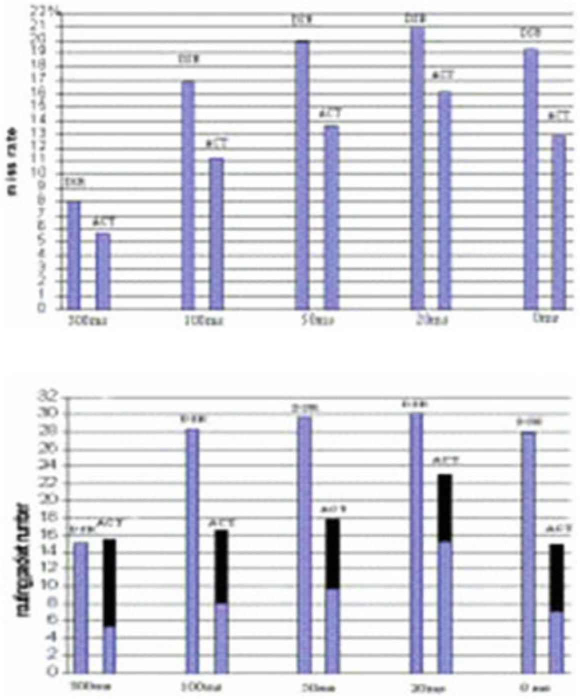


Fig. 4 Miss rates and route packet numbers (per data packet) from the proactive ACT for the 100-node network

References

1. Perkins CE, Royer EM (2009) Ad hoc networking, chapter ad hoc on-demand distance vector routing Addison Wesley, New York
2. Gupta AK (2010) Performance analysis of AODV, DSR and TORA routing protocols. IACSIT int J Eng Technol 2(2):1793–8236
3. Khatri P, Rajput M (2010) Performance study of ad hoc reactive routing protocols. J Comput Sci 6:1150–1154
4. Perkins CE, Bhagwat P (1994) Highly dynamic destination sequenced distance-vector routing (DSDV) for mobile computers. ACM Comput Commun Rev 24(4):234–244
5. Perkins CE, Royer EM (1999) Ad-hoc on-demand distance vector routing, dynamic source routing. In: Proceedings of the 2nd IEEE Workshop on mobile comp systems and applications, New Orlean LA, pp 90–100, Feb 1999
6. Lee A- Xu K (2003) GloMoSim Java Visualization Tool. Documentation version 1.1, Software Distribution
7. IEEE 802.11: part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specification, Aug 1999
8. Stuart K, Tracy C, Michael C (2005) MANET simulation studies: the incredible. Mobile Comput Commun Rev (ACM) 9(4):50–61
9. Johnson D, Maltz D, Hu Y-C, Jetcheva J (2001) The dynamic source routing protocol for mobile ad hoc networks, Internet draft, at http://www.ietf.org/internet-drafts/draft-ietf-manet-dsr-05.txt, March 2001

10. Mahmoud AE, Khalaf R (2007) Kayssi A (2007) Performance comparison of the AODV. DSR and DSDV routing protocols in mobile ad-hoc networks, Lebanon
11. Al- Maashri A, Ould-Khaoua M (2006) Performance analysis of MANET routing protocols in the presence of self- similar traffic. IEEE, ISSN- 0742–1303. First published in proceedings of the 31st IEEE conference on local, computer networks, 2006
12. Mobile Ad Hoc Networks (MANET) (2004) Working Group. http://www.ietf.org/html.characters/manetcharter.html
13. Balakrishna R, Panduranga Rao MV, Shet KC (2008) Development of scheduler for real time and embedded system domain. In: Digital library at IEEE AINA-2008 international conference, JAPAN. The details about the conference is http://www.aina-conference.org/2008
14. Ur Rahman Khan K, Zaman Rafi U, Venugopal Reddy A (2008) Performance comparison of on-demand and table driven ad hoc routing protocols using NCTUns. In: UKSIM: IEEE tenth international conference on computer modeling and simulation, 1(3), pp. 336–341, April 2008
15. Murthy S, Garcia-Luna-Aceves JJ (1996) An efficient routing protocol for wireless networks. ACM Mobile Netw Appl J 1(2):183–197
16. Perkins CE, Royer EM, Das SR (2002) Ad hoc on-demand distance vector (AODV) routing, Internet Draft, draft- ietf- manet aodv-10.txt, work in progress, 2002
17. Bhatt S, Fujimoto R, Ogieski A, Permalla K (1998) Parallel simulation techniques for large scale networks. IEEE Commun Mag 98:42–47

Overhead Analysis of AODV, TORA and AOMDV in MANET Using Various Energy Models

Manish Bhardwaj, Naresh Sharma and Monika Johri

Abstract In this paper we have studied the energy overhead performance of three different routing protocols under three different energy models with some modification. The three different energy models considered are (a) Bansal Energy Model (b) Vaddina Energy Model and (c) Chandrakasan Energy Model. We apply these energy models to AODV, TORA and AOMDV routing protocols to determine the energy overhead among these three routing protocols by varying the transmission range. Our aim is to analyze how these routing protocols behave under different energy models. In the analysis of energy overhead the underlying mobility model also plays a very important role. We have selected the RWP-SS mobility model. In literature many research papers skip the initial simulation time while simulating the routing protocols but this particular mobility model enables us to calculate the energy overhead from the start of the simulation.

Keywords Energy · Overhead · Routing protocols · Simulation · Ad hoc networks · AODV · AOMDV · TORA

1 Introduction

A Mobile Ad hoc Network (MANET) is a kind of wireless ad-hoc network and is a self-configuring network of mobile routers (and associated hosts) connected by wireless links—the union of which forms an arbitrary topology. Energy efficiency plays

M. Bhardwaj (✉) · N. Sharma · M. Johri
Department of Computer Science and Engineering, SRM University, Modinagar, India
e-mail: aapkaapna13@gmail.com

N. Sharma
e-mail: nrssharma@gmail.com

M. Johri
e-mail: johri_monika@yahoo.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_23,

a very important role in the battery operated ad hoc networks. Energy consumption can be due to receiving the data, transmitting the data, traffic, mobility and size of the network. The topology of a mobile ad hoc network changes continuously. Some of the nodes in the path may not be available for transmission leading to broken paths. This results in retransmission of packets. All these activities will deplete the energy available in the nodes very quickly. So, routing algorithms deployed in ad hoc networks play a key role in reducing the overhead involved in energy consumption thus ensuring the longevity of the network.

Most of the research papers proposed in the literature concentrate on the energy consumption and not on the overhead involved. There is a need to study the energy consumption from a routing overhead point of view.

We have made a substantial effort to study the performance of various routing protocols like AOMDV, TORA and AODV under three different energy models like Bansal Energy Model, Vaddina Energy Model and Chandrakasan Energy Model.

The rest of the section is divided as follows. In section two we discuss some of the related work in the literature, we present the different energy models, mobility model and routing protocols in section three, the simulation parameters selected is discussed in four and analysis of result is done in section five and finally we conclude the paper.

2 Related Work

Early work on energy consumption in ad hoc networks was done by Feeney et al. [1]. The authors conduct various experiments to determine the energy consumption of a Lucent Wireless WaveLan IEEE 802.11 network card. The authors also formulate a linear equation to quantify the “per packet energy consumption”.

In [2] the authors have considered various mobility models like Random waypoint (RWP), Manhattan Grid Model (MG), Gauss Markov Model (GM), Community Mobility Model (CM) and RPGM. The authors have analyzed the energy consumption to receive, transmit and drop the control packets. They have calculated the energy consumption by mapping it against the mobility speed. The authors have shown that as the mobility speed is increased the energy output also decreases. Among these three mobility models RWP has the highest energy consumption. For CM and RPGM mobility models it is shown that as the number of groups increases then the energy consumption also decreases.

In [3] the authors have mapped the energy consumption of AODV, DSDV, DSR and TORA routing protocols. The authors have calculated the energy consumed by these four routing protocols by mapping them against varying mobility speed, traffic patterns, node numbers and area. The authors conclude that TORA routing protocol had worst performance in all the scenarios.

In [4] as in [3] three different routing protocols AODV, DSR and DSDV are considered. They are compared against Random Waypoint, RPGM and Manhattan Grid model. They use the same energy model as specified by Feeney. Through simulation the authors show that AODV has more energy consumption under RWP and RPGM, while DSR consumes more energy under Manhattan Grid model.

3 Description of Routing Protocols, Energy Models and Routing Protocols

In this section we give a description of various routing protocols like AODV, TORA, AOMDV, Energy model applied in this paper and RWP-SS mobility model.

3.1 Ad hoc Demand Distance Vector Protocol

The Ad hoc On Demand Distance Vector (AODV) routing algorithm is a routing protocol designed for Ad hoc mobile networks. AODV is capable of both unicast and multicast routing. It is an on demand algorithm, meaning that it builds routes between nodes only as desired by source nodes.

It maintains these routes as long as they are needed by the sources. Additionally, AODV forms trees which connect multicast group members. The trees are composed of the group members and the nodes needed to connect the members. AODV uses sequence numbers to ensure the freshness of routes. It is loop-free, self-starting, and scales to large numbers of mobile nodes. AODV builds routes using a route request/route reply query cycle. When a source node desires a route to a destination for which it does not already have a route, it broadcasts a route request (RREQ) packet across the network. Nodes receiving this packet update their information for the source node and set up backwards pointers to the source node in the route tables. In addition to the source node's IP address, current sequence number, and broadcast ID, the RREQ also contains the most recent sequence number for the destination of which the source node is aware [7]. A node receiving the RREQ may send a route reply (RREP) if it is either the destination or if it has a route to the destination with corresponding sequence number greater than or equal to that contained in the RREQ. If this is the case, it unicast a RREP back to the source. Otherwise, it rebroadcasts the RREQ. Nodes keep track of the RREQ's source IP address and broadcast ID. If they receive a RREQ which they have already processed, they discard the RREQ and do not forward it. As the RREP propagates back to the source, nodes set up forward pointers to the destination. Once the source node receives the RREP, it may begin to forward data packets to the destination. If the source later receives a RREP containing a greater sequence number or contains the same sequence number with a smaller hop count, it may update its routing information for that destination and begin using the better route [6, 7] (Figs. 1 and 2).

AODV is another variant of classical distance vector routing algorithm, based on DSDV and DSR. It shares DSR's on demand characteristics hence discovers routes

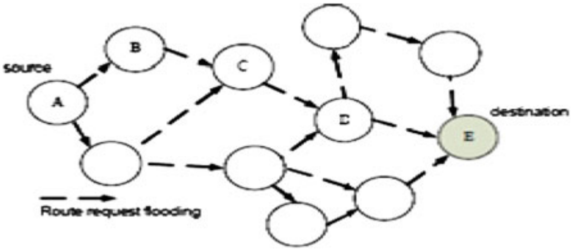


Fig. 1 Route request (RREQ) flooding

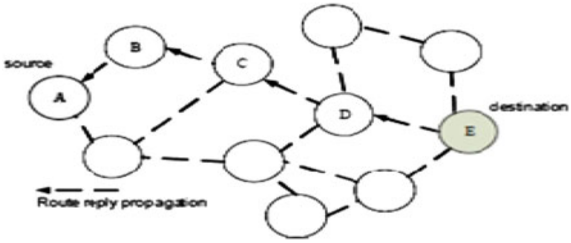


Fig. 2 Route reply (RREP) propagation

whenever it is needed via a similar route discovery process. However, AODV adopts traditional routing table; one entry per destination which is in contrast to DSR that maintains multiple route cache entries for each destination. The initial design of AODV is undertaken after the experience with DSDV routing algorithm [9].
As long as the route remains active, it will continue to be maintained. A route is considered active as long as there are data packets periodically traveling from the source to the destination along that path. Once the source stops sending data packets, the links will time out and eventually be deleted from the intermediate node routing tables. If a link break occurs while the route is active, the node upstream of the break propagates a route error (RERR) message to the source node to inform it of the now unreachable destination(s). After receiving the RERR, if the source node still desires the route, it can reinitiate route discovery [6, 13].

3.2 Temporary Ordered Routing Protocol

TORA is a distributed highly adaptive routing protocol designed to operate in a dynamic multihop network. TORA uses an arbitrary height parameter to determine the direction of link between any two nodes for a given destination. Consequently, multiple routes often exist for a given destination but none of them are necessarily the shortest route.

To initiate a route, the node broadcasts a QUERY packet to its neighbors. This QUERY is rebroadcasted through the network until it reaches the destination or an intermediate node that has a route to the destination. The recipient of the QUERY packet then broadcasts the UPDATE packet which lists its height with respect to the destination. When this packet propagates in the network, each node that receives the UPDATE packet sets its height to a value greater than the height of the neighbor from which the UPDATE was received. This has the effect of creating a series of directed links from the original sender of the QUERY packet to the node that initially generated the UPDATE packet. When it was discovered by a node that the route to a destination is no longer valid, it will adjust its height so that it will be a local maximum with respect to its neighbors and then transmits an UPDATE packet. If the node has no neighbors of finite height with respect to the destination, then the node will attempt to discover a new route as described above. When a node detects a network partition, it will generate a CLEAR packet that results in reset of routing over the ad hoc network.

3.3 Adhoc on Demand Multipath Distance Vector Routing Algorithm

AOMDV employs the "Multiple Loop-Free and Link-Disjoint path" technique. In AOMDV only disjoint nodes are considered in all the paths, thereby achieving path disjointness. For route discovery RouteRequest packets are propagated through out the network thereby establishing multiple paths at destination node and at the intermediate nodes. Multiples Loop-Free paths are achieved using the advertised hop count method at each node. This advertised hop count is required to be maintained at each node in the route table entry. The route entry table at each node also contains a list of next hop along with the corresponding hop counts. Every node maintains an advertised hop count for the destination. Advertised hop count can be defined as the "maximum hop count for all the paths". Route advertisements of the destination are sent using this hop count. An alternate path to the destination is accepted by a node if the hop count is less than the advertised hop count for the destination.

3.4 Random Waypoint Mobility Model–Steady State

While considering the Random Waypoint Mobility Model for simulation, a dissimilar mobility pattern is observed during the initial mobility duration and at the later stage of the simulation. In literature, to avoid the mentioned situation, many of the papers follow a procedure where the initial few seconds are discarded and then it is assumed that the remaining seconds of the simulation are assumed to have a similar pattern. But this method is too crude, as it can not be told at which point the dissimilar pattern starts or stops. To overcome this problem the authors have proposed the Random Way Point-Steady State Mobility Model (RWP-SS). “The initial speed and the stationary distribution location are sampled” to overcome the problem of discarding the initial simulation data. The RWP-SS without pause is given by

$$\mathrm{F}^{-1}(\mathrm{u}) = \mathrm{S}_1^{\mathrm{u}}/\mathrm{S}_0^{\mathrm{u}-1} \tag{1}$$

Here S is the initial speed chosen uniformly over (0, 1) and F-1 (u) is the inverse of the cumulative distribution function. RWP-SS with pause is given by

$$\mathrm{H}_0(\mathrm{p}) = \int_0^{\mathrm{p}} [1 - \mathrm{H}(\mathrm{t})]\mathrm{dt}/\mathrm{E}(\mathrm{p}) \tag{2}$$

where, H(p) is the cumulative distribution function, E(p) is the expected length of a pause.

3.5 Energy Models

For transmission and receiving of energy can be modeled as “E (ptx/rcv) = i * v * t p Joules”, where i is the current value, v is the voltage and tp is the time taken to transmit or receive the packet. The following table gives the various values considered for Bansal Energy Model (BEM), Vaddina Energy Model (VEM) and Chandrakasan Energy Model (CEM).

4 Simulation Environments

Simulations are performed using Network Simulator NS-2 [18]. The simulated values of the radio network interface card are based on the 914 MHz Lucent WaveLan direct sequence spread spectrum radio model. This model has a bit rate of 2 Mbps and a radio transmission range of 250 m. The IEEE 802.11 distributed coordinated function with CSMA/CA is used as the underlying MAC protocol. Interface Queue (IFQ) value of 70 is used to queue the routing and data packets (Tables 1 and 2).

Table 1 Value of different energy models

Energy model	Transmission power	Receiving power power	Idle power
Bansal model	0.0271	0.0135	0.00000739
Vaddina model	0.0744	0.0648	0.00000552
Chandrakasan model	0.175	0.175	0.00000175

Table 2 Various simulation parameters

Parameters	Values
Simulator	NS2
Protocol studied	AODV,TORA,AOMDV
No. of nodes	50
Simulation area	2000x2000
Mobility model	Random way Point SS
Simulation time	500/sec
Transmission range	300,350,400,450,500,550,600,650,700
Maximum speed	10(m/s)
Pause time	10 s
Data payload	512 B
Traffic rate	Packet/sec

5 Result Analyses

Each of the energy overhead models is mapped against the transmission range. The transmission range is varied from 300 to 700 m steps 50. Besides running independently, all the simulations are averaged for 5 different seeds. RWP-SS mobility model enables us to compute the results right from starting of the simulation time. Conserving energy leads to extended battery life in ad hoc networks. Nodes in the ad hoc networks are battery powered. A number of factors shape the extendibility of battery life in ad hoc networks like the speed at which the nodes are moving, the transmission range, the amount of packets sent and received and the amount of information that needs to be processed. It is desirable to find an optimum transmission range to conserve energy without compromising on the amount of data delivered.

Transmission range plays a very important role in deciding the amount of energy overhead needed for establishing connectivity among various nodes in the network. Increasing the transmission range leads to less hop count and there will be less breaks in the connectivity of the mobile nodes.

TORA routing protocol has maximum energy consumption at 400 m across all the energy models. AOMDV has highest energy consumption at 300 m, 350 m and 400 m for BEM, VEM and CEM models. AODV has peak energy consumption at 600 m. Here the idea is not find an optimized transmission or the amount of energy consumed by each protocol. But the intention is to see the behavior of these protocols under various energy models.

AOMDV maintains high connectivity even at high mobility due to multiple paths resulting in less energy overhead for maintaining the network. Even though AOMDV by virtue of its multiple route maintenance has less energy overhead, the maintenance of these multiple paths itself may lead to energy overhead. The amount of energy spent in signaling tends to decrease with increase in transmission range across all the routing protocol. Energy consumption is less in AODV when compared to TORA and AOMDV.

In TORA the number of nodes that can be accessed increases with the increase in transmission range. This increases the amount of interference and collisions, resulting in retransmission of packets. Energy spent in signaling is maximum for TORA protocol at up to 400 m. After this point the energy consumption is comparable to AODV and AOMDV routing protocols. This can be summed to the amount of packets used in maintaining links among various nodes in the network. Establishment of connectivity across the nodes is stabilized with the increase in transmission range. This reduces the energy consumed in the network.

In literature it has been mentioned that the packet delivery increases with increase in transmission range but with higher energy consumption for every transmission. But we found a contrasting result in our simulation. Here energy consumption increases up to a transmission range. After that energy consumption decreases and it remains the same across all the transmission range and then increase slowly again. This tendency can be seen across all the routing protocols.

When the transmission power and receiving power is less then there is huge difference in the amount of energy spent in signaling between AOMDV and AODV routing protocols as can be seen from Figs. 3, 4 and 5. But the difference starts to decrease with the increase in transmission power and receiving power. In CEM model the energy spent in signaling is almost same for both AOMDV and AODV

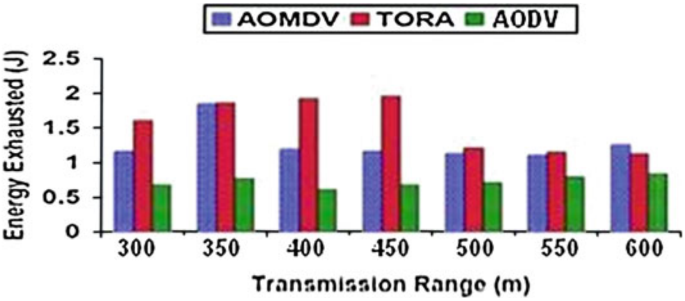


Fig. 3 Energy exhausted (J) v/s transmission range for Bansal energy overhead model

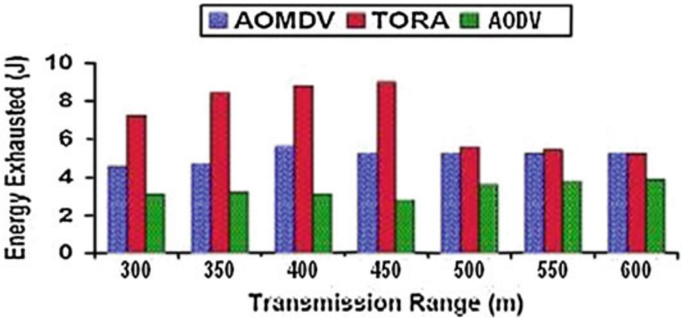


Fig. 4 Energy exhausted (J) v/s transmission range for Vaddina energy overhead model

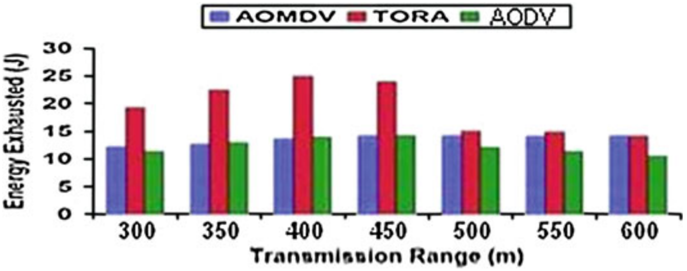


Fig. 5 Energy exhausted (J) v/s transmission range for Chandrakasan energy overhead model

routing protocols. After a transmission range of 400 m the amount of energy spent in signaling between AOMDV and TORA routing protocol remains the same across all the energy models.

6 Conclusions

We have compared the energy overhead involved in various routing protocols by using these energy models. Results are obtained through extensive simulation. We deduce that TORA routing protocol has highest energy overhead across all the three different mobility models. We have also described the role of transmission range in mobile ad hoc networks. For our future work we are planning to investigate the effect of node speed, number of packet sent and received, source and network size on the energy overhead of various other routing protocols.

References

1. Feeney LM, Nilsson M (2001) Investigating the energy consumption of a wireless network interface in an ad hoc netwroking environment
2. Prabhakaran P, Sankar R (2006) Impact of realistic mobility models on wireless networks performance. IEEE international conference on wireless and mobile computing, networking and communications, Montreal, Canada, In, pp 329–334
3. Cano J-C, Manzoni P (2000) A performance comparison of energy consumption for mobile Ad hoc network routing protocols. In: Proceedings of the 8th international symposim on modeling, analysis and simulation of computer and telecommunication systems, IEEE Computer Society, Los Alamitos
4. Chen B-R, Hwa Chang C (2003) Mobility impact on energy conservation of Ad hoc routing protocols. In: SSGRR 2003, Italy
5. Gupta AK (2010) Performance analysis of AODV, DSR and TORA routing protocols. IACSIT Int, J Eng Technol 2:2. ISSN:1793–8236
6. Khatri P, Rajput M (2010) Performance study of Ad hoc reactive routing protocols. J Comput Sci 6:1159–1163

7. Perkins CE, Royer EM (2009) Ad hoc Networking, chapter Ad hoc On-demand distance vector routing. Addison Wesley, Boston
8. Balakrishna R, Panduranga Rao MV, Shet KC (2008) Development of scheduler for real time and embedded system domain. In: IEEE AINA-2008 International conference, JAPAN. http://www.aina-conference.org/2008
9. Rahman Khan K, Rafi Zaman U, Venugopal Reddy A (2008) Performance comparison of on-demand and table driven Ad Hoc routing protocols using NCTUns. UKSIM: IEEE Tenth Int Conf Comput Model Simul 1(3):336–341
10. Mahmoud AE, Khalaf R, Kayssi A (2007) Performance comparison of the AODV and DSDV routing protocols in mobile Ad-hoc networks. Lebanon
11. Ahmed A-M, Mohamed O-K (2006) Performance analysis of MANET routing protocols in the presence of self-similar traffic. In: Proceedings of the 31st IEEE conference on local computer networks. Boston, USA, ISSN-0742-1303
12. Mobile Ad Hoc Networks (MANET) Working Group, http://www.ietf.org/html.characters/manetcharter. Html, 2004
13. Perkins CE, Royer EM, Das SR (2002) Ad hoc on-demand distance vector (AODV) routing. Internet Draft, draft- ietf- manet aodv-10.txt, work in progress
14. Perkins CE, Royer EM (1999) Ad-hoc on-demand distance vector routing. In: Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, pp 90–100.
15. Bhatt S, Fujimoto R, Ogieski A, Permalla K (2002) Parallel simulation techniques for large scale networks. IEEE Comm Mag 98:42–47
16. Lee A, Kaixin X (2004) GloMoSim Java Visualization Tool. Documentation version 1.1, Software Distribution
17. IEEE 802.11: part 11: Wireless LAN Medium Access control (MAC) and Physical Layer (PHY) specification, Aug. 1999.
18. Information Sciences Institute, The Network Simulator Ns-2, http://www.isi.edu/nanam/ns/
19. Stuart K, Tracy C, Michael C (2005) MANET simulation studies: the incredible. ACM Mobile Comput Comm Rev 9:4

Part IV
The Fourth International Conference on Networks & Communications (NETCOM-2012): Network Architectures, Protocols and Routing

Low Power and High Speed Adders in Modified Gated Diffusion Input Technique

R. Uma and P. Dhavachelvan

Abstract Applications of arithmetic operations in integrated circuits are manifold. In most of the digital systems, adders are the fundamental component in the design of application specific integrated circuits like RISC processors, Digital Signal Processors (DSP), microprocessors etc. This paper focuses two main design approaches. The former presents the problems identified in the existing Gate Diffusion Input (GDI) technique and its design solution in Modified Gate Diffusion Input (MGDI). The primitive logic cells are construction in MGDI and its performance issues are compared with existing GDI. The latter presents the implementation of 3 different MGDI full adders and a complete verification and comparison is also carried out to test the performance of the proposed adders. The performance analysis has been evaluated by Tanner simulator using TSMC 0.250 μm technologies. The simulation results reveal better delay and power performance of proposed adder cells as compared to GDI, PT and CMOS at 0.250 μm technologies.

Keywords Full adder · Gate diffusion input · Low power adders · Modified gate diffusion input · Full swing voltage

1 Introduction

In most of the digital systems, adders are the fundamental component in the design of application specific integrated circuits like RISC processors, Digital Signal Processors (DSP), microprocessors etc. The efficient implementation of the addition

R. Uma (✉) · P. Dhavachelvan
Department of Computer Science, School of Engineering, Pondicherry University,
Puducherry, India
e-mail: uma.ramadass1@gmail.com

P. Dhavachelvan
e-mail: dhavachelvan@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_24,

operation in an integrated circuit is a key problem in VLSI design. Many different circuit architectures for binary addition have been proposed [1–11] over the last decades, covering a wide range of performance characteristics. Also, their realization at the transistor level for full-custom circuit implementations has been addressed intensively. Generally, the use of Pass Transistor Logic (PTL) [1] has many advantages over the CMOS design due the reduced transistor count and smaller node capacitances thus decreasing the required area, rise/fall times and power dissipation. However, this scheme suffers from leakage problems because the input inverter is not the full swing signal and it becomes worse when the supply voltage drops. The Complementary Pass-Transistor Logic (CPL) and Swing Restored Pass-Transistor Logic (SRPL) [2–4], resolve the leakage and voltage swing. However, this logic operates in slow speed due to the regenerative property of the latch-type restoring circuit.

The Gate Diffusion Input (GDI) [5, 6] is a lowest power design technique which offers improved logic swing and less static power dissipation. Using this technique several logic functions can be implemented using less number of transistor counts. This method is suitable for design of fast, low-power circuits, using a reduced number of transistors (as compared to TG and CMOS). Though GDI technique offers low power, less transistor count and high speed, the major challenges occurs in the fabrication process. The GDI technique requires twin-well CMOS or Silicon on Insulator (SOI) process to realize a chip which increases the complexity as well as the cost of fabrication.

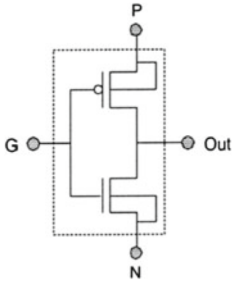
The main contribution of this paper presents the design of modified primitive cells and three different topologies for full adders at circuit level implemented based on the GDI technique. The modified GDI primitive cells are constructed and its significant variation between MGDI and conventional GDI are compared. The organization of the paper is as follows: The Sect. 2, describes the design of modified GDI primitive cells. Section 3, presents the implementation of different MGDI full adders using Tanner EDA. Section 4 presents simulation result using Tanner EDA and it is compared with GDI, CMOS logic and pass transistor logic. Section 5 presents the discussion with different topologies of full adders. Finally the conclusion is presented in Sect. 6.

2 Modified Gate Diffusion Input (MGDI)

The basic structure of GDI consists of N-diffusion, P-diffusion and Gate input. In GDI technique both pMOS and nMOS are given with independent inputs so as to accommodate more logic function thereby minimizing transistor count as well as power dissipation. The main drawbacks associated with GDI are:

- The bulk terminals are not properly biased thereby the circuit exhibits threshold drop and variation in Vt.

Fig. 1 Basic cell in GDI



- Because of floating bulk the cells can be implemented in SOI process which increases the cost of fabrication.
- When N diffusion input is high and P diffusion input is low the diodes between nMOS and pMOS will conduct resulting in static power dissipation. This effect reduces the output voltage to Vout ≈0.5VDD.

The structure of GDI and its logic function is shown in Fig. 1 and Table 1. In the modified Gate Diffusion Input technique the bulk terminals (P-bulk and N-bulk) are permanently tied to VDD and GND thereby the threshold drop variation can be surmounted. The other advantage of this configuration provides suitability for fabricating the logic cells in CMOS p-well and n-well process. The other modification in MGDI is, expect for inverter, function F1 and F2 all other cells are realized with gate input thereby static power dissipation can be reduced. To obtain the full swing voltage the output of each cell has between incorporated with buffers. The MGDI and its functional table are shown in Fig. 2 and Table 2. Figure 3 shows the construction

Table 1 Logic function implemented with GDI technique

N	P	G	OUT	Function
'0'	*B*	*A*	$\overline{A}B$	*F1*
B	*'1'*	*A*	$\overline{A}+B$	*F2*
'1'	*B*	*A*	$A+B$	*OR*
B	*'0'*	*A*	AB	*AND*
C	*B*	*A*	$\overline{A}B+AC$	MUX
'0'	*'1'*	*A*	$\overline{A}$	*NOT*

Fig. 2 Basic cell in MGDI

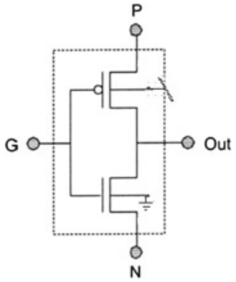


Table 2 Logic function implemented with MGDI technique

N	P	G	Out	Function
0	B	A	$\overline{A}$B	F1
B	1	A	$\overline{A}$+B	F2
A	B	A	A+B	OR
B	A	A	AB	AND
B	$\overline{A}$	A	$\overline{AB}$	NAND
$\overline{A}$	B	A	$\overline{A+B}$	NOR
C	B	A	$\overline{A}$B+AC	MUX
0	1	A	$\overline{A}$	NOT
B	$\overline{B}$	A	$\overline{A}$B+A$\overline{\mathbf{B}}$	XOR
$\overline{B}$	B	A	$\overline{A}\overline{B}$+A$\mathbf{B}$	XNOR

of modified basic gates of AND, NOR, OR, NAND, XOR, XNOR and MUX in MGDI. The proposed MGDI primitive cells are simulated using Tanner EDA with BSIM3v3 250 nm technology with supply voltage ranging from 0 to 5 V in steps of 0.2 V. The performance of GDI and MGDI is presented in the Table 3.

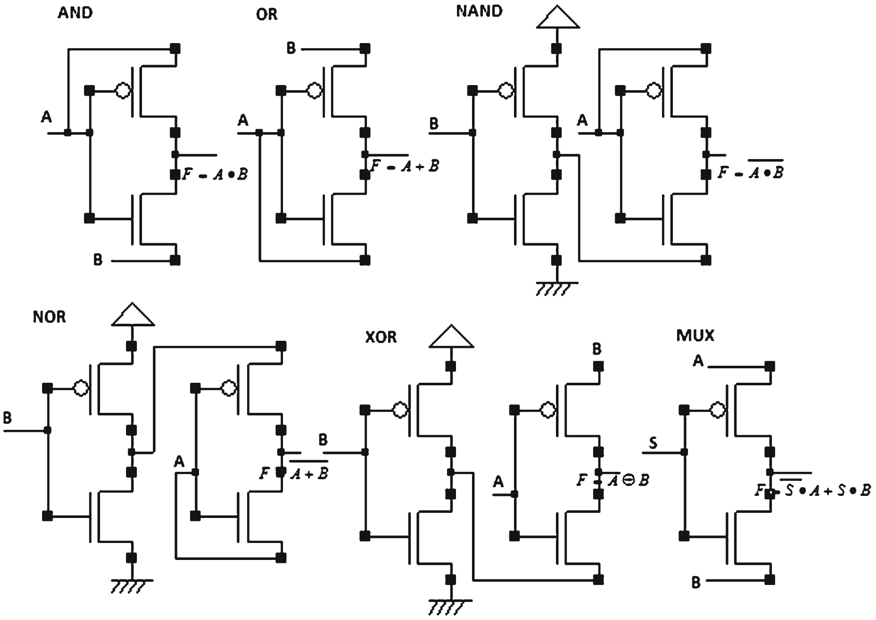

AND
A
B
$F = A \bullet B$
OR
B
A
$F = A + B$
NAND
B
A
$F = \overline{A \bullet B}$
NOR
B
A
F $\overline{A+B}$
XOR
B
B
A
F $\overline{A} \ominus B$
MUX
A
S
B
$F = \overline{S} \bullet A + S \bullet B$

Fig. 3 Proposed modified GDI primitive gates

Table 3 Performance of GDI and MGDI

Gate type	Delay (ns)	Rise delay (ns)	Fall delay (ns)	Power	Tr count	AT	AT^2	PD
NAND2GDI	0.42758	2.276	2.0305	100.1	4	1.71032	0.731299	42.80076
NAND2MGDI	0.4821	2.0818	2.3673	98.1	6	2.8926	1.394522	37.29401
NAND3GDI	0.7834	3.1266	3.1456	110.1	10	7.834	6.137156	86.25234
NAND3MGDI	0.8121	2.1235	2.3334	100.2	14	11.3694	9.23309	71.37242
NAND4GDI	1.2451	5.6123	5.5431	120.34	16	19.9216	24.80438	149.8353
NAND4MGDI	1.3567	4.2312	4.2562	105.23	22	29.8474	40.49397	132.7655
NOR2GDI	2.51	2.003	2.1256	48.1	4	10.04	25.2004	120.731
NOR2MGDI	2.61	1.1256	1.3256	45.32	6	15.66	40.8726	108.2852
NOR3GDI	4.221	4.003	4.1256	65.78	10	42.21	178.1684	277.6574
NOR3MGDI	4.678	2.546	2.6568	52.34	14	65.492	306.3716	244.8465
NOR4GDI	6.221	6.1233	6.5786	93.12	16	99.536	619.2135	602.4805
NOR4MGDI	6.978	4.1686	4.7868	86.34	22	153.516	1071.235	579.2995
AND2GDI	0.53715	3.9603	3.2914	100.32	6	3.2229	1.731181	55.88689
AND2MGDI	0.5673	3.1155	3.1234	98.45	8	4.5384	2.574634	43.85069
AND3GDI	0.8795	5.1223	5.1231	115.34	12	10.554	9.282243	101.4415
AND3MGDI	0.9123	4.2812	4.2162	105.23	16	14.5968	13.31666	86.00133
AND4GDI	1.9571	6.1723	6.5431	122.46	18	35.2278	68.94433	239.6665
AND4MGDI	1.9567	5.2312	5.2562	110.23	24	46.9608	91.8882	205.687
OR2GDI	2.61	2.1233	2.2231	49.12	6	15.66	40.8726	128.2032
OR2MGDI	2.783	1.563	1.4578	45.33	8	22.264	61.96071	126.1534
OR3GDI	4.345	3.4563	3.3431	52.512	12	52.14	226.5483	228.1646
OR3MGDI	4.523	3.1	3.1271	45.33	16	72.368	327.3205	205.0276
OR4GDI	6.1845	5.5676	5.4531	68.785	18	111.321	688.4647	425.4008
OR4MGDI	6.562	4.5745	4.6782	55.874	24	157.488	1033.436	366.6452
XORGDI	27.346	9.5635	9.5123	110.34	4	109.384	2991.215	3017.358
XNORGDI	27.346	9.5635	9.5123	110.34	4	109.384	2991.215	

3 Implementation of Low Power Full Adders in MGDI

A full adder is a combinational circuit that performs the arithmetic sum of three bits: A, B and a carry in, C, from a previous addition produces the corresponding SUM, S, and a carry out, CARRY. The various equations for SUM and CARRY are given below

Adder1

SUM=$A \oplus B \oplus C$ (1)

CARYY=$(\overline{A \oplus B})$A+$(\overline{A \oplus B})$C (2)

Adder3

SUM=$\overline{A \oplus B \oplus C}$ (5)

CARYY=$(\overline{A \oplus B})$A+$(A \oplus B)$C (6)

Adder2

SUM=$\overline{C}(A \oplus B)$+ C(A$\oplus$ B) (3)

CARRY=$(\overline{A \oplus B})$A+$(A \oplus B)$C (4)

Three different MGDI full adders have been designed with transistor count of 10 T. The circuit is implemented with and without restoration circuit.

The maximum output produced for the circuit implemented without restoration is 4.3 V with the input voltage of 5 V. The circuit included with buffer produces full swing output. Adder1 is implemented with 3-input XOR for sum logic and 2-to-1 MUX for carry logic. The transistor count for this circuit with and without buffer is 14 T and 10 T. Adder2 is designed from the Eqs. 3 and 4, the sum expression is designed using 2-input XOR gate, NOT and 2-to-1 MUX, whereas carry expression is designed using 2-to-1 MUX with transistor count of 18 and 10 T. Similarly adder3 is implemented with 3-input XNOR for sum logic and 2-to-1MUX for carry logic with transistor count of 14 and 10 T. The design of 3 different full adders using MGDI with and without restoration is shown in Fig. 4.

4 Simulation and Performance Analysis

In order to test the performance of the proposed MGDI full adders, detailed comparisons are performed. The proposed designs are simulated using TSMC 0.250 μm CMOS process. The cell is supplied with supply voltage ranging from 1 to 5 V in steps of 0.5 V. The three inputs to the full adder are A, B, C and all the test vectors are generated and have been fed into the adder cell. The cell delay has been measured from the moment the inputs reach 50 % of the voltage supply level to the moment the latest of the SUM and CARRY signals reach the same voltage level. All transitions from an input combination to another (total 8 patterns 000, 001, 010, 011, 100, 101, 110, 111) have been tested, and the delay at each transition has been measured. The simulation setup is shown in Fig. 5. Figures 6 and 7 show the output waveform of MGDI 10T full adder and primitive cells.

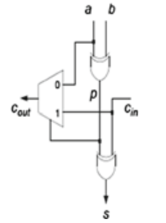
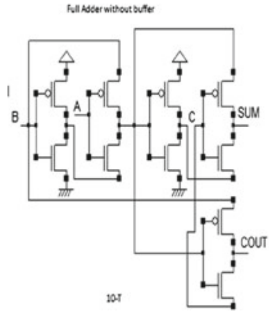
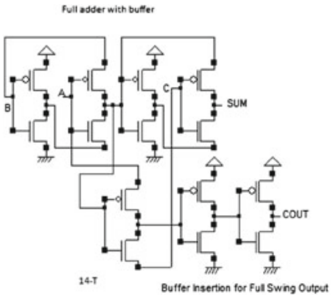
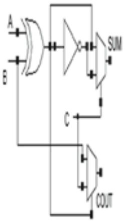
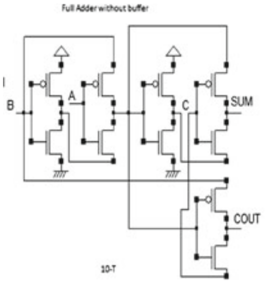
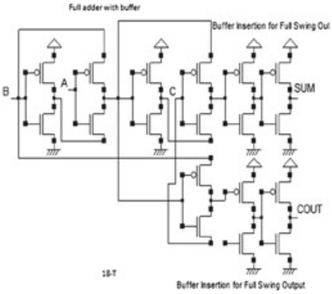
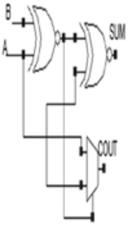
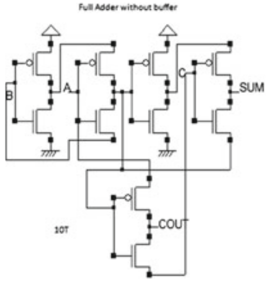
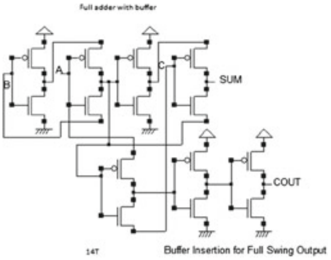
Adder	Logic diagram	Full adder without Buffer	Full adder with Buffer
ADDER 1		 Full Adder without buffer I B A C SUM COUT 10-T	 Full adder with buffer B A C SUM COUT 14-T Buffer Insertion for Full Swing Output
ADDER 2		 Full Adder without buffer I B A C SUM COUT 10-T	 Full adder with buffer Buffer Insertion for Full Swing Out B A C SUM COUT 18-T Buffer Insertion for Full Swing Output
ADDER 3		 Full Adder without buffer B A C SUM COUT 10T	 Full adder with buffer B A C SUM COUT 14T Buffer Insertion for Full Swing Output

Fig. 4 Proposed MGDI full adders

The performances of different MGDI full adders with and without restoration circuit have been analyzed in terms of delay, transistor count and power dissipation with respect to CMOS, pass transistor logic and GDI technique.

It is observed that modified GDI full adder cells have least delay and power consumption when compared to CMOS, pass transistor logic and GDI technique. Table 4 presents the performance of MGDI, GDI, CMOS and pass transistor.

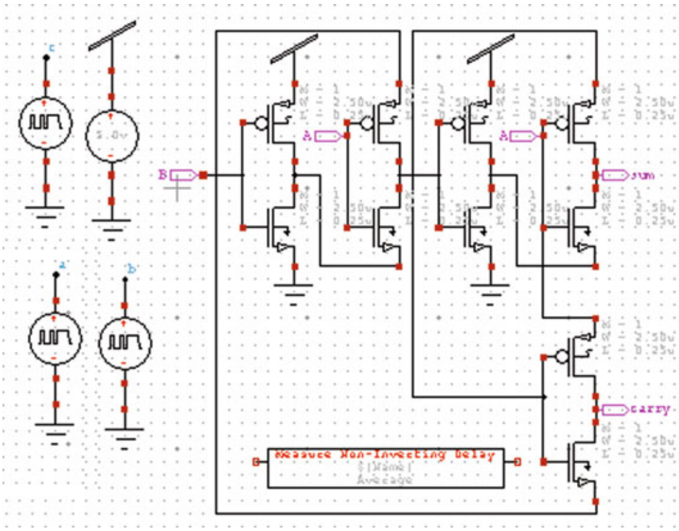


Fig. 5 Simulation setup and test for MGDI 10 T full adder circuit

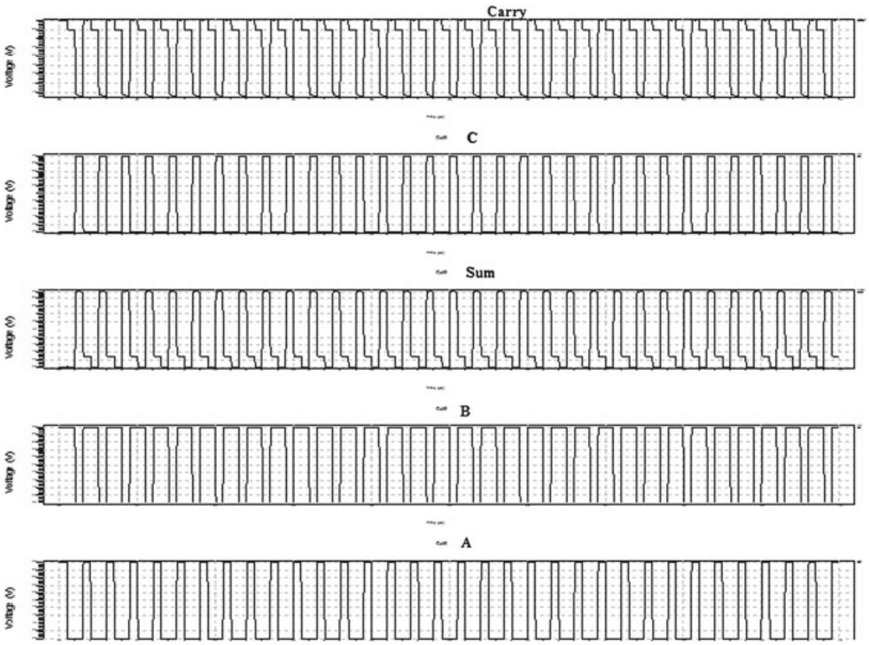


Fig. 6 Input/output waveform of MGDI 10 T full adder circuit

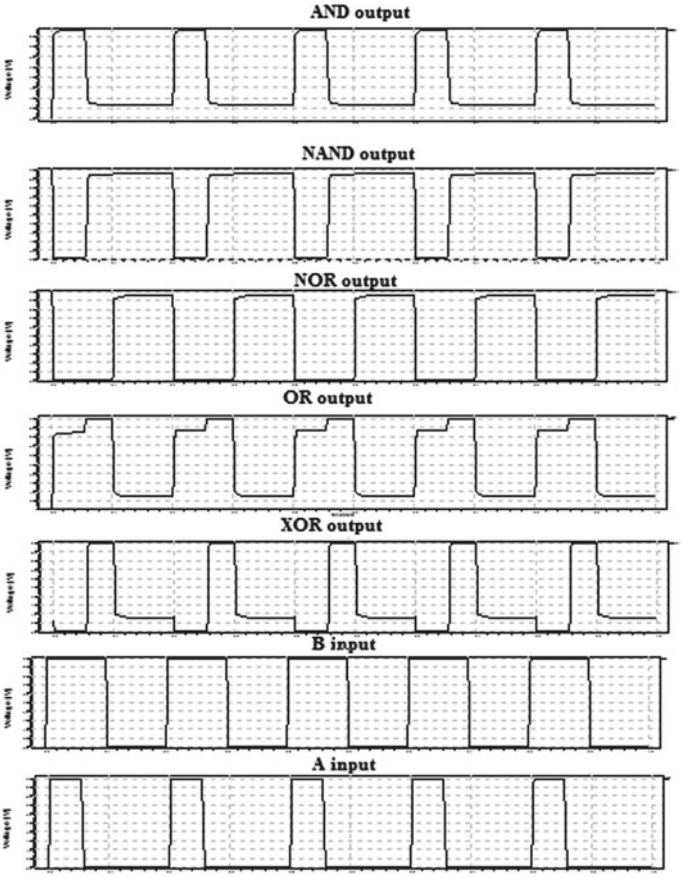


Fig. 7 Input/output waveform of MGDI primitive cells

Table 4 Performance of GDI and MGDI

	Delay in (ps)				Transistor count				Power dissipation (μW)			
Full adder type	MGDI	GDI	CMOS	PT	MGDI	GDI	CMOS	PT	MGDI	GDI	CMOS	PT
Adder1	32.3 32.9	33.1	35.1	34.1	10 14	18	38	22	12.1 10.2	12.5	12.8	15.7
Adder2	29.1 29.5	29.7	30.3	29.8	10 18	16	36	20	10.6 9.6	10.1	12.7	14.9
Adder3	26.3 26.7	27.2	29.1	28.3	10 14	14	30	18	8.34 7.89	8.99	10.7	13.5

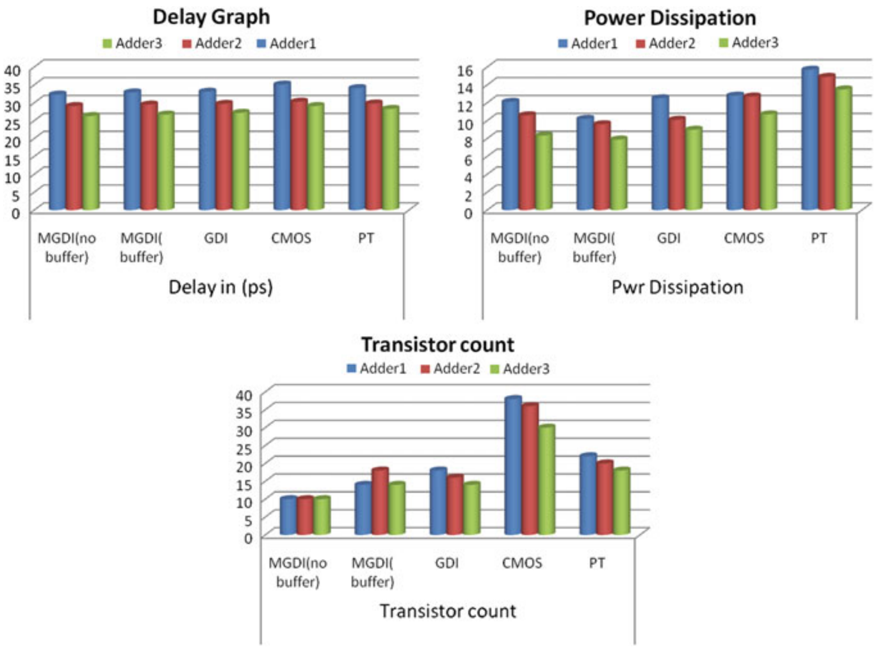


Fig. 8 Comparison of full adders in terms of delay, transistor count and power dissipation

5 Discussion

Three different full adder cells are proposed using XOR, XNOR and MUX primitives. Its performances have been analyzed in MGDI with and without restoration circuit, GDI, CMOS and Pass transistor logic are reported in Sect. 4. The comparison of delay, transistor count and power dissipation is depicted in Fig. 8.

From the graph it can be observed that adder3 constructed with XNOR and MUX primitives exhibits the least delay and power dissipation. It is also noticed that the full adder designed with MGDI with buffers produces less power dissipation and delay when compare to MGDI without buffers.

6 Conclusion

An extensive performance analysis of modified primitive cells of AND, OR, NAND, NOR, MUX, XOR and XNOR has been presented. The performance of these MGDI was analyzed in terms of transistor count, delay and power dissipation using Tanner EDA with TSMC MOSIS 250 nm technology. Subsequently three different 1-bit full-adder cells have been presented and their analyses were also reported. From this analysis it can be observed that adder3 constructed with XNOR and MUX primitives exhibits the least delay and power dissipation. It is also noticed that the full adder designed with MGDI with buffers produces less power dissipation and delay when compare to MGDI without buffers, GDI, CMOS and PT.

References

1. Chatzigeorgiou A, Nikolaidis S (2001) Modelling the operation of pass transistor and CPL gates. Int J Electron 88(9):977–1000
2. Yano K, Yamanaka T, Nishida T, Saito M, Shimohigash K, Shimizu A (1990) A 3.8-ns CMOS 16×16 multiplier using complementary pass-transistor logic. IEEE J Solid State Circ 25:388–394
3. Parameswar A, Hara H, Sakurai T (1994) A high speed, low power, swing restored pass-transistor logic based multiply and accumulate circuit for multimedia applications. Custom integrated circuits conference, In, pp 278–281
4. Yano K, Sasaki Y, Rikino K, Seki K (1996) Top-down pass-transistor logic design. IEEE J Solid State Circ 31:792–803
5. Morgenshtein A, Fish A, Wagner IA (2002) Gate-diffusion input (GDI): a power-efficient method for digital combinatorial circuits. IEEE Trans Very Large Scale Integr Syst 10(5):566–581
6. Uma R, Dhavachelvan P (2012) Performance of adders with logical optimization in FPGA. First international conference on signal, image processing and pattern recognition (SPPR 2012) May 2012, Delhi, published in advances in intelligent and soft computing book series of springer, signal, image processing and pattern, pp 245–255
7. Uma R, Dhavachelvan P (2012) ASIC primitive cells in modified gated diffusion input technique. In: International conference on advances in computing (ICADC 2012) July 2012. Bangalore, Published in the book series of Springer, Advances in Intelligent Systems and Computing, Vol 174:467–474
8. Uma R, Dhavachelvan P, (2012) Performance of full adder with skewed logic. In: International Conference on Advances in Computing and Communications (ICACC-2012) August 2012 Kochin, Published by IEEE Computer Society, pp 150–153
9. Ng P, Balsara PT, Steiss D (1996) Performance of CMOS differential circuits. IEEE J Solid State Circ 31(6):841–846
10. Aguirre-Hernandez M, Linares-Aranda M (2011) CMOS full-adders for energy-efficient arithmetic applications. IEEE Trans Very Large Scale Integr Syst 19(4):525–537
11. Uma R, Dhavachelvan P (2012) Analysis on impact of behavioral modeling in performance of syn thesis process. In: The third international workshop on VLSI (VLSI 2012) July 2012. Chennai, Published in the book series of Springer, Advances in Intelligent and Computing, Vol 178:593–602

Guided Local Search for Optimal GPON/FTTP Network Design

Ali Rais Shaghaghi, Tim Glover, Michael Kampouridis and Edward Tsang

Abstract Fibre optic access networks are of increased popularity among network operators. Many types of fibre network are being deployed globally to satisfy the ever increasing users bandwidth requirements. The rate of deployments of such networks is expected to increase in coming years, moreover this requires cost efficient, reliable and robust network designs. Despite the relative complex structure of these networks, designs are mostly done manually, thus design quality is not optimal. In this paper we will introduce and propose a tree based modelling scheme and will show how the metaheuristic search method Guided Local Search can be used to automate the design of FTTP/GPON networks. The design optimisation will mainly focus on reducing the deployment cost i.e finding the optimal location, type and quantity of fibre optic equipment in order to reduce the capital expenditure (CAPEX) of such deployment projects. Our proposed model builds a flexible optimisation framework, and results of the GLS algorithm compared to simple local search and Simulated Annealing show consistent optimal results.

Keywords Guided local search · Fibre optic networks · Network optimisation · Network planning

1 Introduction

With the growth of telecommunication networks and the introduction of new applications and services, the demand for higher bandwidths is increasing rapidly. With these increases of bandwidth demands fibre optic networks are becoming the

A. R. Shaghaghi (✉)
CCFEA, University of Essex, Colchester, UK
e-mail: araiss@essex.ac.uk

T. Glover
BT Research and Technology, Ipswich, UK

M. Kampouridis, E. Tsang
CSEE, University of Essex, Colchester, UK

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_25,

preferred solution. Hybrid networks with mixed use of fibre optics and copper cables have been around for a while, however they have limited capacity. For this reason operators are moving towards FTTP (Fibre-To-The-Premises) technologies, based on the Gigabit Passive Optical Network (GPON), which enables them to provide services that are ready to meet customers demand with higher bandwidths. FTTP lines are projected to almost triple globally, from 68 million in 2011 to 198 million in 2016, driven by uptake in China, Russia, and the US and increasing deployments from Western European incumbent operators. Pyramid Research expects FTTH broadband to generate nearly $ 116 billion in service revenue by 2016 worldwide, creating opportunities for all of the stakeholders in the value chain.[1] The United Kingdom is showcasing its broadband initiative, and FTTH deployment in most of England is growing rapidly as well.[2]

However, the deployment and further costs associated with FTTx.[3] Networks are comparatively greater than legacy copper access networks [1]. The competitive service-providing market motivates the network operators to design and deploy more economic networks with relativity low capital expenditure (CAPEX). Their target is to bring their FTTP costs down as close to the legacy copper telecommunication networks as possible.

One of the key aspects in reducing deployment cost is to have an efficient low cost design, meaning for a given deployment plan reducing various equipment and labour costs associated with design.

A cost efficient design depends on positioning optical components in the underlying road and duct network, so as to minimise the number of components and the length of fibre cable required. In addition to these there are several constraints that have to be satisfied in the design and the planning of the network, such as the maximum outputs any specific splitter can have. All these considerations require tools that provide efficient and robust network designs and deployment plans. In a typical GPON deployment, an exchange area is often divided into different sections each served by an aggregation node. Each aggregation node can be seen as root to a tree of splitters, fibre distribution points (FDPs) and manifolds (see Fig. 1). Once the location of this equipment is determined then the planner has to design the layout of cables from each manifold to FDPs and each FDP to a splitter respectively. These planning tasks are time consuming and the efficiency of the design and the associated costs are dependent on the planner's experience because of the simple fact that they are done manually. In this paper we propose a tree-based model to represent the network design problem and will specifically introduce a customised Guided Local Search algorithm [2] to achieve an efficient design. The combination of the

[1] http://www.prnewswire.com/news-releases/ftth-lines-expected-to-triple-by-2016-finds-pyramid-research-135128728.html

[2] http://www.lightwaveonline.com/business/news/Frost–Sullivan-FTTH-deployments-lend-momentum-to-European-fiber-optic-test-equipment-markets-124229344.html

[3] Fiber to the x (FTTx) is a generic term for any type of access broadband network which uses fibre optic as its main transmission medium, all starting by FTT, these variations are differentiated based on different configurations (e.g. FTTN, FTTP, FTTH, and so on, where in the above examples "N" denotes Node, "C" denotes Premises, "H" denotes Home).

network representation model and the metaheuristic optimisation method will result in an automated tool that enables network operators to efficiently plan and design robust and efficient GPON/FTTP networks. The automated tool will generate solutions that are considered optimal or near optimal with respect to cost and satisfying specific design constraints. The introduction of automated planning tool in context of GPON/FTTP networks results in these advantages: (1) Rapid generation of network design layouts. (2) Support for exploring different scenarios by changing constraints (3) Minimising CAPEX (4) Automatically producing bills of material (5) Providing detailed implementation costs for techno-economic analysis.

The rest of the paper is organised as follows: Sect. 3 will introduce a tree-based representation model to solve the optimisation problem followed by description of the metaheuristic algorithm for design automation in Sect. 4. In Sect. 5 we describe our experimental design and results and Sect. 6 will provide future works and conclusion.

2 Related Work

There have been several studies on optimising various types of telecommunication networks some of which, are specifically related to fibre optic network designs.

In [3] the authors have proposed a model to optimise the design of a GPON/FTTH network, their model considers certain green field design aspects and a mixed integer linear programming solver is used to find a near optimal solution. Their solution promises a design with satisfactory degree of symmetry in addition to short computational time.

In [4] they propose an efficient heuristic called the Recursive Association and Relocation Algorithm (RARA) to solve the optimization problem. Their model also propose splitting large areas into smaller optimisation problems in order to reduce the computation time.

The other proposed model in [4] describes an algorithm that recursively assigns network elements to the design layout, their research provides a good theoretical lower bound on the deployment cost for PON networks. For more complex design cases that consider constraints such as road maps and other geographic constraints, sub-optimal solutions can be extended from their planning approaches.

3 Model Description

When installing a new network in the access area, the majority of money has to be spent on digging the cable ducts. Thus, minimizing the total cost is mainly a matter of finding the shortest street paths which interconnect all optical network units (ONUs)[4]

[4] In our model we will call these points Exchange, Aggregation, FDP (Fibre Distribution Point) and Splitter Nodes.

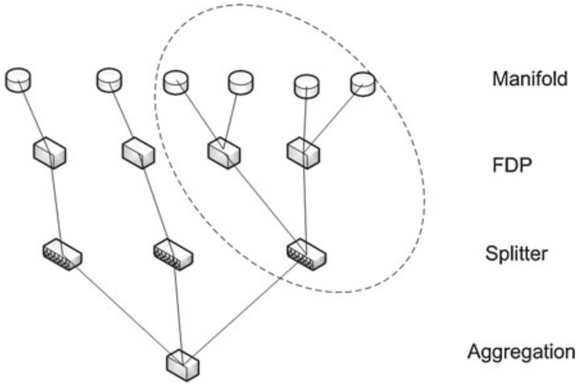


Fig. 1 Logical configuration

with the optical line termination (OLT).[5] A city map can be represented by a graph where the streets are the links, and the street junctions together with the ONUs and the OLT make up the nodes. The weights of the links are set to be proportional to the length of the respective streets. In some cases, for example, if some fibre lines exist or if some streets are preferred to be used as duct lines, special weight values can be assigned to theses edges. With this map representation, the optimization problem turns into the classical minimum Steiner tree problem. This means that we want to find a tree within a given graph which spans a designated subset of nodes in such a way that the sum of the costs of the selected edges becomes minimal. There already exists a number of algorithms that solve this problem exactly. Since the minimum Steiner tree problem is NPcomplete, these algorithms have an exponential worst-case running time. Therefore, they are not applicable in the field of network planning where it is quite common to have a great number of nodes and edges [5].

The representation has two parts. The first maps optical components to geographical locations. Each piece of equipment is represented by a variable whose domain ranges over possible locations. By restricting the domain we can constrain the equipment to a limited geographical area.

The second part of the model describes how the optical components are connected to each other. The optical tree is considered as a collection of clients and servers. For example each splitter serves many FDPs, and each FDP could be served by one of many splitters. Moving an FDP from one splitter to one another changes the connectivity of the tree. This one to many relationship is represented by including a variable for each client whose domain ranges over its possible servers. Each component may act as both client and sever for example a splitter has FDPs as its clients and an aggregation node as its server .

Figure 1 shows one possible logical configuration and Fig. 2 shows how part of this network can be laid out on a physical road network. The cables will take the shortest path between connected points. All clients of the server passed en route to

[5] Known as Manifolds in our model.

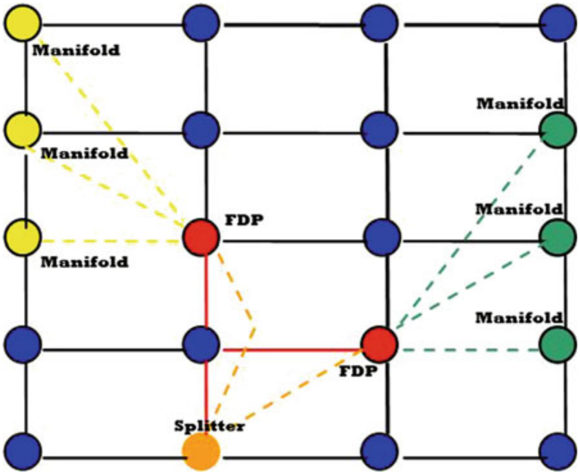


Fig. 2 Physical layout

the most distant client will be attached to the same cable. The cost of the network is the cost of components plus the cost of the cables.

Design constraints are controlled by adding penalties to solutions that violate the constraints. Design constraints include for example the capacity of junction boxes, maximum length of cables, branch aggregation factor of splitters and so on. The total cost to be minimised by optimisation process is : TotalCost = Componants Cost + Cabling Cost + Penalty Cost

4 Local Search

In order to find an optimal solution to this problem we will use Local Search. Local Search encompasses a class of algorithms which move from solution to solution in the search space of candidate solutions by applying local changes until no further improvements can be found or until a time limit has been exceeded. In our problem representation a move is an assignment of a different value to one or more variables. Table 1 represents the implemented moves in this local search.

Table 1 Local search moves

Move types	Description
Assignment	Assign a new value to a single variable
Swap	Swap the value of two variables
Join domains	Moves all the clients of a server to another server
Split domain	Moves half of the clients of one server to a different server

In this section we will introduce two variants of local search; Hill Climbing and Simulated Annealing (SA), we then describe the metaheuristic, Guided Local Search (GLS).

4.1 Hill Climbing and Simulated Annealing

Our proposed metaheuristic will sit on top of a tailored local search scheme designed for our proposed network optimisation model. The local search simply model solution initialization, new solution generation (neighbourhood function), and improved solution acceptance.In Hill Climbing a move to a new solution is only accepted if it results in an improved cost. Therefore there is a monotonic improvement in cost. The disadvantage of this approach is that it is unable to escape from a local minimum. Simulated Annealing attempts to overcome this by allowing an "uphill" move with a probability that decreases over time. The allowance for "uphill" moves potentially saves the method from becoming stuck at local optima. In our simple configuration of SA there are 1000 iterations with the probability of move acceptance exponentially decaying with a rate of 0.9.

4.2 Guided Local Search

Most of local search methods suffer from two disadvantages. Firstly they easily get stuck in local minima. Secondly, in many cases we have intuition about how to guide the search but this can not be included directly in the cost function. For example, in the Travelling Salesman Problem, we know that long edges are undesirable though we can not exclude them from the beginning because they may be needed to connect remote clusters of cities in the problem. Guided Local Search (GLS) is a penalty-based approach that sits on top of local search methods which can help solve these problems. When the given local search algorithm is trapped in a local optimum, GLS dynamically changes the objective function, by penalizing some selected features that are present in this solution. This raises the cost of the solution, allowing the search to continue. The features are chosen in such a way as to guide the search towards promising areas by giving an incentive to remove unfavourable features. The novelty of GLS is mainly in the way that it selects problem dependent features to penalize, determined by two factors: the feature's cost (i.e. influence on the objective function) and the frequency with which it has been penalised in the search so far [2]. These features should simply satisfy the constraint of being non trivial, meaning that they would not appear in all solutions [6]. If S is a set of all possible solutions the presence of a feature f_i in solution $s \in S$ is represented by an indicator function

$$I_i(s) = \begin{cases} 1 & s\ has\ feature\ f_i \\ 0 & otherwise \end{cases}$$

Associated with each feature f_i is a cost c_i, and a penalty p_i which counts the number of times this feature has been penalised. When a local minimum is reached a feature is chosen to be penalised by a utility function which considers the cost of the feature and its current penalty. The utility function is defined by $util(s, f_i) = I_i(s) \cdot \frac{c_i}{1+p_i}$. Features which have already been penalised are less likely to be penalised again. This reflects the intuition that we should avoid selecting the same feature every time. Augmenting the cost function g via penalties on features gives us a new objective function h defined : $h(s) = g(s)+\lambda\cdot\sum_{i=1}^{M} p_i \cdot I_i(s)$ where M is the number of features defined over solutions and λ is a regularization parameter. A number of different possible features were explored for this problem and the most effective was found to be pairs of consecutive items of equipment on a cable. The feature cost is the cost of cable linking the two items.

5 Experimental Design and Results

The solver has to find a solution that connects all the manifolds to the exchange whilst satisfying all the problem constraints and minimising the cost via reducing the cabling and total equipment cost.

The selected region includes one aggregation node and one exchange, the final solution will layout cables from the single exchange point to 85 Manifolds. The constraints shown in Table 2 describe the maximum number of connections that the equipment could have. The number of children in this table indicates the maximum number of clients that can be served by each item of optical equipment. Also there is a upper bound limit for the possible number of FDPs and Splitters, which are 43 and 11 respectively.

In order to evaluate the effectiveness of GLS two solvers were compared, the first using simulated annealing and the second using Guided Local Search with features based on cables as described earlier. Each experiment was run 50 times over the sample experimental data to find the optimal solution. In each case the search was allowed to continues until no improvement had been found for more than 6 minutes.

Table 2 Equipment constraints

	Connections	Children		Connections	Children
Manifold	12	0	Exchange	100000	100
Aggregation	276	20	Splitter	4	24
FDP	24	3			

Table 3 Best cost statistics for 50 runs

	GLS	SA		GLS	SA
Mean	13958.72	14263.19	Skewness	1.015195	2.249336
Standard deviation	297.7969	690.6522	Min	13621.93	13613.12
Kurtosis	0.06251	6.583823	Max	14742.49	16977.58

Given equal execution time we are interested in the most optimal cost (minimised) that derives from the automated design of the network. Table 3 depicts the statistics for SA and GLS. The results simply imply more consistent performance from GLS algorithm in comparison to higher standard deviation of the SA. The average cost of the network is also smaller while using GLS. For the sake of statistical analyses we have performed a two sample t-test with the null hypothesis that data in the vectors of SA and GLS are independent random samples from normal distributions with equal means and equal but unknown variances, against the alternative that the means are not equal. The results shown in Table 3 allow us to reject the null hypothesis.

For further proof of the effectiveness of our metaheuristic methods we have also tested a simple hill climbing algorithm,[6] the results obtained simply shows very poor performance i.e. the solver in this case failed to satisfy many of the problems constrains failing to produce any acceptable solution. The nature of GLS algorithm enables it to scape settlements in local minima therefore results prove to be more consistent.

6 Conclusion

Designing fibre optic access networks is becoming of great interest to global telecom providers. In this paper we have presented an automated GPON/FTTP design framework based on a tree-based model utilising a guided local search algorithm to find a near optimal solution. The model structure is relatively flexible enabling production of various network designs with different constraints and requirements. The automated algorithm enables network designers and planners to quickly plan GPON networks with high flexibility and near optimal solutions. We use Guided Local Search (GLS) to eliminate the common problem of local search algorithms getting trapped in local optimum solutions. The GLS metaheuristic tends to produce robust results in many runs thus ensuring rapid solutions with high quality. The ability of GLS to escape local minima provides significantly higher quality results.

[6] This iterative method starts with an arbitrary solution and tries to find a better solution by using the described local search moves. If the change results in better solutions it will accept the solution until no further improvements occur

References

1. Verbrugge S, Casier K, Lannoo B, Van Ooteghem J, Meersman R, Colle D, Demeester P (2008) FTTH deployment and its impact on network maintenance and repair costs. In: 10th anniversary international conference on transparent optical networks ICTON (2008) vol 3 . IEEE, New York, pp 2–5
2. Voudouris C (1999) Guided local search and its application to the traveling salesman problem. Eur J Oper Res 113(2):469–499
3. Ouali A, Poon K (2011) Optimal design of GPON/FTTH networks using mixed integer linear programming. In: 16th European conference on networks and optical communications (NOC), IEEE, pp 137–140
4. Li J, Shen G (2008) Cost minimization planning for passive optical networks. In: OFC/NFOEC 2008–2008 conference on optical fiber communication/national fiber optic engineers conference, pp 1–3, Feb 2008
5. Riedl A (1998) A versatile genetic algorithm for network planning. In: Proceedings of EUNICE, vol 98, Citeseer, pp 97–103
6. Voudouris C, Tsang E (2003) Guided local search. Handbook of metaheuristics. In: Glover F (ed) Handbook of Metaheuristics. Kluwer, Dordrecht, pp 185–218

Image Segmentation Using Variable Kernel Fuzzy C Means (VKFCM) Clustering on Modified Level Set Method

Tara Saikumar, Khaja FasiUddin, B. Venkata Reddy and Md. Ameen Uddin

Abstract In this paper, Variable Kernel Fuzzy C-Means (VKFCM) was used to generate an initial contour curve which overcomes leaking at the boundary during the curve propagation. Firstly, VKFCM algorithm computes the fuzzy membership values for each pixel. On the basis of VKFCM the edge indicator function was redefined. Using the edge indicator function the image segmentation of a medical image was performed to extract the regions of interest for further processing. The above process of segmentation showed a considerable improvement in the evolution of the level set function.

Keywords Image segmentation · VKFCM · Level set method

1 Introduction

Image segmentation is plays an important role in the field of image understanding, image analysis, pattern identification. The foremost essential goal of the segmentation process is to partition an image into regions that are homogeneous (uniform) with respect to one or more self characteristics and features. Clustering has long been a popular approach to un tested pattern recognition. The fuzzy c-means (FCM) [1] algorithm, as a typical clustering algorithm, has been utilized in a wide range of engineering and scientific disciplines such as medicine imaging, bioinformatics, pattern recognition, and data mining. *Given* a data $X = \{x_i \ldots x_n\} \subset R^p$, the original FCM

T. Saikumar (✉) · Md. Ameen Uddin
Department of ECE, CMR Technical Campus, Hyderabad, Andhra Pradesh, India
e-mail: tara.sai437@gmail.com

K. FasiUddin
Department of ECE, VITS, Karimnagar, Andhra Pradesh, India

B. V. Reddy
Department of ECE, GNITS, Hyderabad, Andhra Pradesh, India

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_26,

algorithm partitions X into c fuzzy subsets by minimizing the following objective function

$$J_m(U, V) \equiv \sum_{i-1'}^{c} \cdot \sum_{k-1}^{n} u_{ik}^{m} \|x_i - v_i\|^2 \tag{1}$$

where c is the number of cluster and selected as a specified Value in the paper, n the number of data points, u_k, the member of x_k in class i , satisfying $\sum_{i-1}^{c} u_{ik}$, m the quantity controlling clustering fuzziness and v is set of control cluster centers or a prototypes $(v_i \in R^p)$. The function J_m is minimized by the famous alternate iterative algorithm. Since the original FCM uses the squared-norm to measure inner product with an appropriate 'kernel' function, one similarity between prototypes and data points, it can only be effective in clustering 'spherical' clusters. And many algorithms are resulting from the FCM in order to cluster more general dataset. Most of those algorithms are realized by replacing the squared-norm in Eq. (1) the object function of FCM with other similarity trial (metric) [1, 2]. In this paper, a variable kernel based fuzzy c-means algorithm (VKFCM) is projected. VKFCM adopt a new kernel-induced metric in the data space to restore the original Euclidean norm metric in FCM. By replacing the inner product with an appropriate 'kernel' function, one can absolutely perform a nonlinear mapping to a high dimensional feature space without increasing the number of parameters.

The level set method is [3–6] based on geometric deformable model, which translate the problem of evolution 2-D (3-D) close curve(surface) into the evolution of level set function in the space with higher dimension to obtain the advantage in managing the topology changing of the shape. The level set method has had great success in computer graphics and vision. Also, it has been widely used in medical imaging for segmentation and shape recovery [7, 8]. However, there are some insufficiencies in traditional level set method.

Finally, if the initial evolution contour is given at will, the iteration time would increase greatly; too large or too small contour will cause the convergence of evolution curve to the contour of object incorrectly. Therefore, some modification has been proposed to improve the speed function of curve evolution [9–11]. In the paper, based on the new variational level set method, the edge indicator function was weighted to improve the ability of detecting fuzzy boundaries of the object. At the same time, the VKFCM algorithm [12, 13] was applied to obtain the appropriate initial contour of evolution curve, so as to get the accurate contour of object and reduce the evolution time.

2 Variable Kernel Fuzzy C-Means Clustering (VKFCM)

Define a nonlinear map as $\phi : x \rightarrow \phi(x) \in F$, where $x \in X$. X denotes the data space and F is the transformed feature space with higher even infinite dimensions. VKFCM minimized the following objective function:

$$J_m(U, V) \equiv \sum_{i-1'}^{c} \cdot \sum_{k-1}^{n} u_{ik}^{m} \|\phi(x_i) - \phi(v_i)\|^2 \tag{2}$$

where

$$\|\phi(x_i) - \phi(v_i)\|^2 = K(x_k, x_k) + K(v_i, v_i) - 2K(x_k, v_i) \tag{3}$$

where $K(x, y) = \phi(x)^T \phi(y)$ is an inner product of the kernel function. If we adopt the Gaussian function as a kernel function, $K(x, y) = \exp(-\|x - y\|^2/2\sigma^2)$, then $K(x, x) = 1$. According to Eqs. (3), (2) can be rewritten as

$$J_m(U, V) \equiv 2 \sum_{i-1'}^{c} \cdot \sum_{k-1}^{n} u_{ik}^{m} (1 - k(x_k, v_i)). \tag{4}$$

Minimizing Eq. (4) under the constraint of, $u_{ik}, m > 1$. We have

$$u_{ik} = \left[\frac{(1/(1 - K(x_k, v_i)))^{1/(m-1)}}{\sum_{j=1}^{c} (1/(1 - K(x_k, v_i)))^{1/(m-1)}} \right]^{1/2} \tag{5}$$

$$v_i = \left[\frac{\sum_{k=1}^{n} u_{ik} K(x_k, v_i) x_k}{\sum_{k=1}^{n} u_{ik}^{m} K(x_k, v_i)} \right]^{n} \tag{6}$$

Here we now utilize the Gaussian kernel function for Straightforwardness. If we use additional kernel functions, there will be corresponding modifications in Eqs. (5) and (6).

In fact, Eq. (3) can be analyzed as kernel-induced new metric in the data space, which is defined as the following

$$d(x, y) \underline{\underline{\Delta}} \|\phi(x) - \phi(y)\| = \sqrt{2(1 - K(x, y))} \tag{7}$$

And it can be proven that $d(x, y)$ is defined in Eq. (7) is a metric in the original space in case that $K(x, y)$ takes as the Gaussian kernel function. According to Eq. (6), the data point x_k is capable with an additional weight $K(x_k, v_i)$, which measures the similarity between x_k and v_i and when x_k is an outlier i.e., x_k is far from the other data points, then $K(x_k, v_i)$ will be very small, so the weighted sum of data points shall be more strong.

The full explanation of VKFCM algorithm is as follows:

KFCM Algorithm:

Step 1: Select initial class prototype $\{v_i\}_{i=1}^{c}$.

Step 2: Update all memberships $u_{ik}$with Eq. (5).

Step 3: Obtain the prototype of clusters in the forms of weighted average with Eq. (6).

Step 4: Repeat step 2–3 till termination. The termination criterion is $\| V_{new} - V_{old} \| \leq \varepsilon$.

where $\|.\|$ is the Euclidean norm. V is the vector of cluster centers ε is a small number that can be set by user (here $\varepsilon = 0.01$).

3 The Modification to the Level Set Method

The level set method was invented by Osher and Sethian [3, 14] to hold the topology changes of curves. A simple representation is that when a surface intersects with the zero plane to give the curve when this surface changes, and the curve changes according with the surface changes. The heart of the level set method is the implicit representation of the interface. To get an equation describing varying of the curve or the front with time, we started with the zero level set function at the front as follows:

$$\phi(x, y, t) = 0, \text{if}(x, y) \in 1 \tag{8}$$

Then computed its derivative which is also equal to zero

$$\frac{\partial \phi}{\partial t} + \frac{\partial \phi}{\partial x} . \frac{\partial x}{\partial t} + \frac{\partial \phi}{\partial y} . \frac{\partial y}{\partial t} = 0 \tag{9}$$

Converting the terms to the dot product form of the gradient vector and the x and y derivatives vector, we go

$$\frac{\partial \phi}{\partial t} + \left(\frac{\partial \phi}{\partial x} . \frac{\partial x}{\partial t} \right) . \left(\frac{\partial \phi}{\partial y} . \frac{\partial y}{\partial t} \right) = 0 \tag{10}$$

Multiplying and dividing by $\nabla \phi$ and taking the other part to be F the equation was gotten as follows:

$$\frac{\partial \phi}{\partial t} + F|\nabla \phi| = 0 \tag{11}$$

According to literature [8, 10], an energy function was defined:

$$E(\phi) = \mu E_{\text{int}}(\phi) + E_{ext}(\phi) \tag{12}$$

where $E_{ext}(\phi)$ was called the external energy, and $E_{\text{int}}(\phi)$ was called the internal energy. These energy functions were represented as:

$$E_{\text{int}}(\phi) = \int\limits_{\Omega} \frac{1}{2}(\nabla\phi - 1)^2 dxdy \tag{13}$$

$$E_{ext}(\phi) = \lambda L_g(\phi) + vA_g(\phi) \tag{14}$$

$$L_g = \int\limits_{\Omega} g\delta(\phi)|\nabla\phi| dxdy \tag{15}$$

$$A_g = \int\limits_{\Omega} gH(-\phi)dxdy \tag{16}$$

$$g = \frac{1}{1 + |\nabla G_\sigma * I|} \tag{17}$$

where $L_g(\phi)$ was the length of zero level curve of ϕ; and A_g could be viewed as the weighted area; I was the image and g was the edge indicator function. In conventional(traditional) level set methods, it is numerically necessary to keep the evolving level set function close to a signed distance function [13, 15]. Re-initialization, a technique for occasionally re-initializing the level set function to a signed distance function during the evolution, has been extensively used as a numerical remedy for maintaining stable curve evolution and ensuring desirable results.

From the practical viewpoints, the re-initialization process can be quite convoluted, expensive, and has subtle side effects [16]. In order to overcome the problem, Li et al. [7, 10] proposed a new variational level set formulation, which could be easily implemented by simple finite difference scheme, without the need of re-initialization. The details of the algorithm are in the literature [7, 10]. However, because only the gradient information was imposed in the edge indicator function, Li's method has a little effect on the presence of fuzzy boundaries.

In the paper, a innovative method was proposed to modify the algorithm. The original image was partitioned into some sub images by VKFCM. The fuzzy boundary of each sub image was weighted by α, the edge indicator function was redefined:

$$g' = g + \alpha \cdot g_2 \tag{18}$$

where $g_2 = \dfrac{1}{1 + |\nabla G_\sigma * I_1|}$

I_1 Was the image after clustering. The iterative equation of level set functional was:

$$\frac{(\phi^{n+1} - \phi^n)}{\tau} = \mu\left[\Delta\phi - div\left(\frac{\nabla\phi}{|\nabla\phi|}\right)\right] + \lambda\delta(\phi)div\left(g'\frac{\nabla\phi}{|\nabla\phi|}\right) + vg'\delta(\phi) \tag{19}$$

Taking $g' = g + \alpha \cdot g_2$ into 19

$$\phi^{n+1} = \phi^n + \tau \left\{ \begin{array}{l} \mu \left[\nabla\phi - div \left(\frac{\nabla\phi}{|\nabla\phi|} \right) \right] \lambda\delta\left(\phi\right) div \left(g \frac{\nabla\phi}{|\nabla\phi|} \right) \\ +vg\delta\left(\phi\right) + \alpha \left[\begin{array}{l} \lambda\delta\left(\phi\right) div \left(g_2 \frac{\nabla\phi}{|\nabla\phi|} \right) + \\ vg_2\left(\phi\right) \end{array} \right] \end{array} \right\} \tag{20}$$

where $\alpha \in [0, 1]$. When processing images of weak boundary or low contrasts, a bigger α was taken; otherwise, a smaller $\alpha$was taken.

4 The Generation of Initial Contour Curve

On the basis of VKFCM clustering in image segmentation, the over segmentation usually exists. In this paper, the result of VKFCM was used as initial contour curve, and the automated initialization of deformation model was finished. For all the pixels in each cluster i.e. white matter, if 4 neighborhoods included the heterogeneous pixel, the pixel was regarded as candidate boundary point. So the algorithm of curve tracing [17–19] was proposed. The exterior boundary of the cluster was tracked in the candidate boundary points. The steps of image segmentation with adapted level set method were as follows:

Step1. Set the number of clusters, then the original image was processed with VKFCM, and calculate the g_2.

Step2. Choose one cluster, evaluate the inside area with $-\rho$ and the outside area with $+\rho$, ρ is a plus constant. The boundary of the area is set to 0. The region of interest is defined initial contour.

Step3. Minimize the overall energy functional with 20 formula.

5 Experimental Results

The segmentation of image takes an important branch in the surgery navigation and tumor radiotherapy. However, due to medical imaging characteristics, the low contrast and fuzzy boundary is usually occurred in the images. In the experiment, the samples of images are taken from internet as shown in figure.

With the enhanced the approximate contour of white matter was got by VKFCM algorithm shown in Fig. 1b. The snooping of regions else appear as a result of the in excess of segmentation.

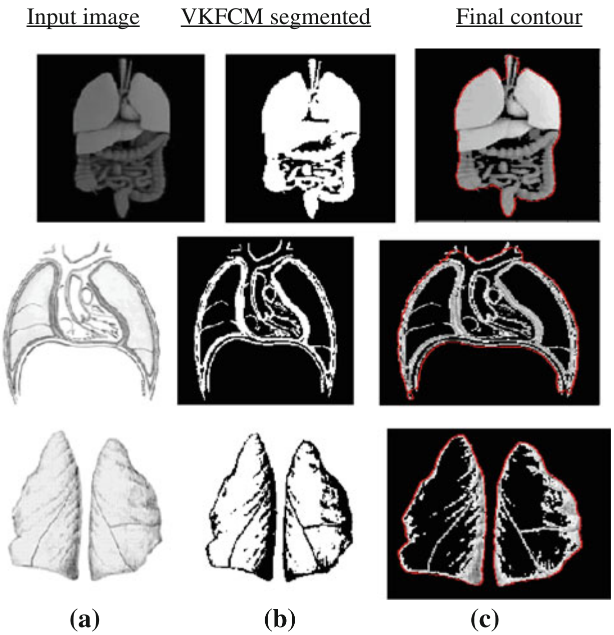


Fig. 1 **a** Original test images, **b** Results of VKFCM clustering, to extracting the *white* matter. **c** Results of final contour with proposed method

6 Discussion

The need of the re-initialization is completely eliminated by the proposal of Chunming Li, for pure partial differential equation driven level set methods, the variational level set methods. It can be easily implemented by using simple finite difference method and is computationally more efficient than the traditional level set methods. But, in this algorithm, the edge indicator has little effect on the low contrast image. So it is hard to obtain a perfect result when the region has a fuzzy or discrete boundary. Meanwhile, the initial contour of evolution needs to be determined by manual, and it has the shortcomings of time consuming and user intervention.

In this paper, we projected a new method to transform the algorithm. The original image was partitioned with VKFCM, and the controlled action of the edge indicator function was increased. The result of VKFCM segmentation was used to obtain the initial contour of level set method. With the new edge indicator function, results of image segmentation showed that the improved algorithm can exactly extract the corresponding region of interest. Under the same computing proposal, the average time cost was lower. Alternatively the VKFCM clustering is sensitive to noise; some redundant boundaries were appeared in the candidates. Consecutively to solve this problem, the algorithm of curve tracing was proposed.

7 Conclusion

In conclusion, the results of this study confirmed that the combination of Improved KFCM with the level set methods could be used for the segmentation of low contrast images. The method has the advantages of no reinitialization, automation, and reducing the number of iterations. The validity of new algorithm was verified in the process of exacting different images. In the future research, the effect of priori information on the object boundary extraction with level set method, such as boundary, shape, and size, would be further analyzed. At the same time, the performance of image segmentation algorithms would be improved by reconstruction of classic velocity of level set method.

References

1. Bezdek JC (1981) Pattern recognition with fuzzy objective function algorthims. Plenum Press, New York
2. Wu KL, Yang MS (2002) Alternative c-means clustering algorithms. Pattern Recognit 35(10):2267–2278
3. Osher S, Sethian JA (1988) Fronts propagating with curvature dependent speed: algorthim's based on the Hamilton-Jacobi formulation. J Comput Phys 79(1):12–49
4. Malladi R, Sethain J, Vemuri B (1995) Shape modeling with front propagation: a level set approach. IEEE Trans Pattern Anal Mach Intell 17(2):158–175
5. Staib L, Zeng X, Schultz R, Duncan J (2000) Shape constraints in deformable models. In: Bankman IN (ed) Handbook of medical imaging. Academic Press, New York, pp 147–157
6. Leventon M, Faugeraus O, Grimson W (2000) Wells W (2000) Level set based segmentation with intensity and curvature priors. Workshop on mathematical methods in biomedical image analysis proceedings, In, pp 4–11
7. Paragios N, Deriche R (2000) Geodesic active contours and level sets for the detection and tracking of moving objects. IEEE Trans pattern Anal Mach Intell 22:266–280
8. Vese LA, Chan TF (2002) A multiphase level set frame wor for image segmentation using the mumford and shah model. Int J Comput Vis 50(3):271–293
9. Shi Y, Karl WC (2005) Real-time tracking using level set. In: Proceedings of IEEE computer society conference on computer vision and, pattern recognition, vol 2, pp 34–42.
10. Li C, Xu C, Gui C et al (2005) Level set evolution without re-initialization: a new varitional formulation. IEEE computer society conference on computer vision and pattern recognition, In, pp 430–436
11. Sethain I (1999) Level set methods and fast marching methods. Cambridge University Press, Cambridge
12. Dunn JC (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J Cybern 3:32–57
13. Bezedek J (1980) A convergence thheorem for the fuzzy ISODATA clustering algorthims. IEEE Trans Pattern Anal Mach Intell 2:78–82
14. Zhang L, Zhou WD, Jiao LC (2002) Kernel clustering algorithm (in chinese). Chin J Comput 25(6):587–590
15. Osher S, Fedkiw R (2002) Level set methods and dynamic implicit surfaces. Springer, New York, pp 112–113
16. Peng D, Merrimam B, Osher S, Zhao H, Kang M (1996) A PDE-based fast local level set method. J Comp Phys 155:410–438

17. Gomes J, Faugeras O (2000) Reconciling distance functions and level sets. J Vis Commun Image Represent 11(2):209–223
18. Mcinerney T, Terzopouls D (1996) Deformable models in medical image analysis: a survey. Med Image Anal 1(2):9l–108
19. Dao-Qiang Z, Song C (2003) Clustering in completed data usig Kernel-based fuzzy c-means algorthim. Neural Process Lett 18(3):155–162

Comparison of SPIHT, Classical and Adaptive Lifting Scheme for Compression of Satellite Imageries

K. Nagamani, A. G. Ananth and K. V. S. Ananda Babu

Abstract The signals that are encountered in practice are not smooth signals and classical wavelet transforms cannot deal with the discontinuities encountered in the signals. Such singularities tend to give rise to large coefficients in their proximity, which is undesirable for signal compression. To overcome such problems one can consider the local variance during the decomposition of the signal. There are various ways to build adaptivity into the decomposition of a signal. The best algorithm, selects a wavelet basis by minimizing a concave cost function such as the entropy. In such an approach, the filter coefficients are fixed for entire block of data as the optimization criterion is global. Here, the decompositions are considered where the filter coefficients vary locally, taking into account of local signal variations. The approach taken by Chan and Zhou [1] suggests that instead of changing the filter coefficients, the input signal is changed in the proximity of discontinuities through an extrapolation procedure. By registering these changes, the original signal can be recovered at synthesis level. By extending the approach of Chan and Zhou in the present work, the SPIHT, Classical Lifting scheme and Adaptive Lifting schemes are analyzed for achieving better compression ratio and PSNR for satellite Rural and Urban imageries. The results are presented in the paper.

Keywords Set partitioning in hierarchical trees (SPIHT) · Discrete wavelet transform (DWT) · Peak signal to noise ratio (PSNR) · Compression ratio (CR) · Cohen-Daubechies-Feauveau CDF

K. Nagamani (✉) · A. G. Ananth
Department of Telecommunication Engineering, R.V. College of Engineering,
Bangalore, India
e-mail: knmsm_03@yahoo.com

K. V. S. Ananda Babu
C M R Institute of Technology, Bangalore, India
e-mail: antisro@yahoo.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_27,

1 Introduction

Over the past decade wavelet-based image compression schemes have become increasingly important and gained widespread acceptance. Because of their inherent multiresolution signal representation, wavelet-based coding schemes have the potential to support both SNR and spatial scalability.

An improved scheme, called Set Partitioning in Hierarchical Trees (SPIHT), was developed by Said and Pearlman. It employs a manner in which all the insignificant coefficients are placed in larger subsets, thus reducing the necessity of transmitting separate bits for every insignificant sub-trees.

Lifting scheme was first introduced by Sweldens. The lifting scheme is a new method for constructing bi-orthogonal wavelets (Sweldens 1996b). The basic idea behind the lifting scheme is that, it gradually builds a new wavelet, corresponding to resized version of the image with improved properties, by adding new basis functions [1–7].

Trappe and Liu build adaptivity into the prediction step of the lifting scheme. Their aim is to design a data-dependent prediction filter to minimize the predicted detail signal. The filter coefficients at a given time are updated using the approximation signal and the predicted detail at time. In this scheme, perfect reconstruction is automatic [1–7].

2 Lifting Scheme Forward Transform

Lifting scheme forward transform consists of three steps: 1. Split 2. Predict 3. Update.

Split: In split step the data is divided into ODD and EVEN elements.

Predict step: The difference between the odd and the even data forms the odd elements of the next step wavelet transformation. The predict step, where the odd value of next iteration is "predicted" from the even value of present step is described by the Eq. 1. Index 'j' represents iteration and 'i' represents element:

$$odd_{j+1,\,i} = odd_{j,\,i} - P\left(even_{j,\,i}\right) \tag{1}$$

Update step: The update step replaces the even elements of the next step with an average of the earlier step. These results in a smoother input (even element) for the next step wavelet transform. The update step follows the predict phase. So in calculating an average the update phase must operate on the differences that are stored in the odd elements

$$even_{j+1,\,i} = even_{j,j} + U\left(odd_{j+1,\,i}\right) \tag{2}$$

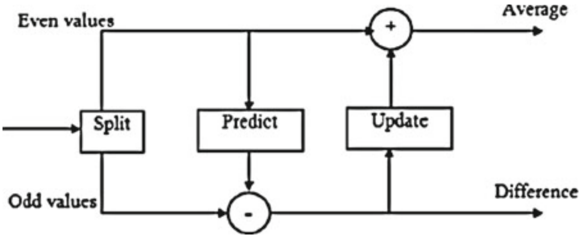


Fig. 1 Lifting scheme forward transform

A simple lifting scheme forward transform is shown in Fig. 1.

After dividing the complete data set into two parts that is even and odd, the processing is done as follows: For the forward transform iteration j and element i, the new odd element j + 1, i would be

$$odd_{j+1,i} = odd_{j,j} - even_{j,i} \tag{3}$$

Even element of the next step is calculated as the original value of the $odd_{j+1,i}$ element has been replaced by the difference between this element and its even predecessor. Simple algebra recovers the original value from Eq. 3.

$$odd_{j,i} = even_{j,i} + odd_{j+1,i} \tag{4}$$

Substituting this into the average that is Eq. 4 we get

$$even_{j+1,i} = \frac{even_{j,i} + odd_{j,i} + odd_{j+1,i}}{2} \tag{5}$$

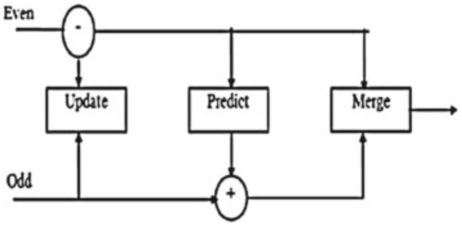
$$even_{j+1,i} = even_{j,i} + \frac{odd_{j+1,i}}{2} \tag{6}$$

The averages (even elements) become the input for the next recursive step of the forward transform. The number of data elements processed by the wavelet transform must be a power of two. If there are 2n data elements, the first step of the forward transform will produce 2n−1 averages and 2n−1 differences (between the prediction and the actual odd element value). These differences are sometimes referred to as wavelet coefficients.

2.1 Lifting Scheme Inverse Transform

Inverse lifting scheme transform as the name suggested is the mirror process of forward lifting scheme transform. We recover the original data sequence by going upwards as shown in Fig. 2 above. Additions are substituted for subtractions and subtractions for additions. The merge step replaces the split step.

Fig. 2 Lifting scheme inverse transform



3 Adaptivity in Wavelet Transforms

The lifting predictors work well when the underlying signal is smooth. However, most of the images consist of regions of smoothness and texture separated by discontinuities (edges). These discontinuities cannot be well-represented by smooth basis functions. Since smooth basis functions correspond to lifting predictors with wide support, these predictors work poorly near edges, when the discontinuity is within the data that are used for prediction.

The introduction of a mechanism that allows to choose the prediction operator based on the local properties of the image. This makes the P operator data-dependent and thus the transform is nonlinear. However, lifting guarantees that the transform remains reversible. In regions where the image is locally smooth, it uses higher order predictors. Near edges it reduces the order and thus the length of the predictor. Such an adaptation would allow us to exploit the spatial structure that exists in edges.

3.1 Adaptive Update Lifting Scheme

An edge detection algorithm analyzes the data in the 2-D prediction window to determine the location and the orientation of the edge. When an edge pixel is detected then we use a lower order predictor. In the study an edge detection algorithm using Sobel operator is considered (Gonzalez et al. 2004). The Sobel operator performs a 2-D spatial gradient measurement on an image. Typically it is used to find the approximate gradient magnitude at each point in an input grayscale image. The classical operator such as Sobel, which uses first derivative, has very simple calculation to detect the edges and their orientations. It is easy to implement than the other operators. Sobel operator effectively highlights noise found in real world pictures as edges though the detected edges could be thick. Hence, Sobel operator is highly recommended in massive data communication found in image data transfer (Hafiz et al. 2011).

The Sobel edge detector uses a pair of 3 × 3 convolution masks, one estimating the gradient in the x-direction and the other estimating the gradient in the y-direction. A convolution mask is usually much smaller than the actual image. As a result, the mask is slid over the image, manipulating a square of pixels at a time.

In lossy compression the decoder has only the quantized even coefficients rather than the original coefficients. If we use locally adapted filters, then quantization errors in coarse scales could cascade across scale and cause a series of incorrect filter choices leading to serious reconstruction errors. The simple modification that solves this problem is to reverse the order of the predict and update lifting steps in the wavelet transform as shown in Fig. 3.

4 Results and Discussion

The SPIHT, Adaptive update lifting and Classical Lifting algorithm has been implemented on MATLAB software and tested for three sample gray images of size 256 × 256 with 8 bit per pixel. For the satellite Rural image, satellite Urban image, and the standard Lena image. The decomposition level is varied 1–5. The corresponding values of MSE, PSNR, and Rate distortion are determined for the three types of images and for different levels of decomposition for SPIHT, Lifting and Adaptive Lifting Scheme.

It is evident from the Fig. 4 that for the SPIHT scheme the PSNR values increases rapidly with increase of Decomposition levels from 1 to 5 and reaches a maximum PSNR value for the 5th level of Decomposition. This suggests that the PSNR values

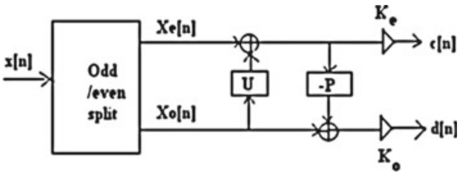
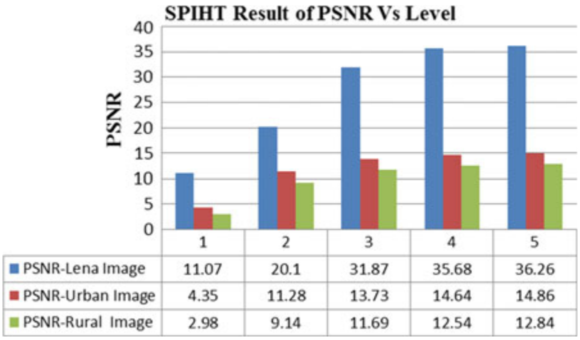


Fig. 3 Update first lifting scheme



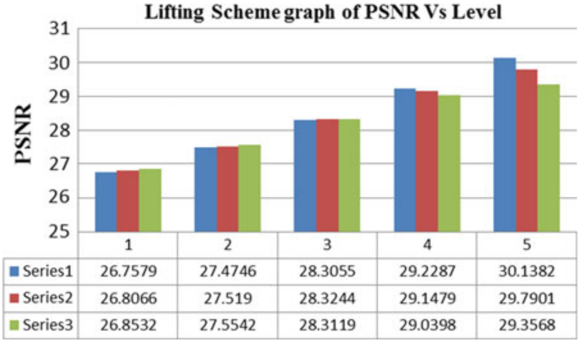
	1	2	3	4	5
🟦 PSNR-Lena Image	11.07	20.1	31.87	35.68	36.26
🟥 PSNR-Urban Image	4.35	11.28	13.73	14.64	14.86
🟩 PSNR-Rural Image	2.98	9.14	11.69	12.54	12.84

Fig. 4 SPIHT results PSNR versus decomposition level

are very low for the Rural and Urban Imageries and SPIHT scheme is not very efficient for compression of satellite imageries.

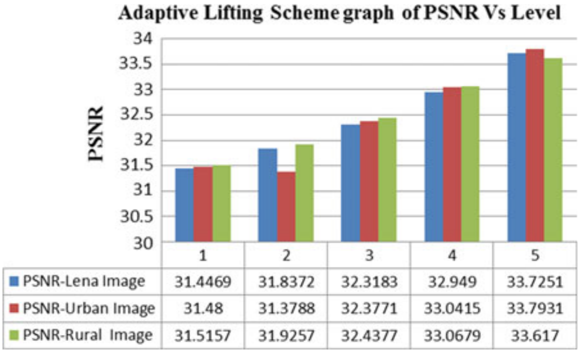
It may be seen from Fig. 5 that for the Classical Lifting SPIHT scheme, the PSNR values increases slowly with increase of Decomposition levels from 1 to 5 and reaches a maximum PSNR value for the 5th level of Decomposition. The PSNR values are very high for the Rural and Urban Imageries compared to SPIHT scheme. The lifting scheme is found almost independent of decomposition levels and very efficient for compression of satellite imageries.

The Fig. 6 depicts the PSNR values for the Adaptive Lifting SPIHT scheme. It can be noted that for all the three images, the PSNR values increases slowly with increase of decomposition levels from 1 to 5 and reaches a maximum PSNR value for the 5th level of decomposition. The adaptive lifting scheme the three images shows further improvement in PSNR to classical lifting scheme. The PSNR values are found to be very high for the Rural and Urban Imageries compared to SPIHT scheme and classical lifting scheme. The adaptive lifting scheme the PSNR is found almost independent of decomposition levels and highly efficient for compression of satellite imageries.



	1	2	3	4	5
🟦 Series1	26.7579	27.4746	28.3055	29.2287	30.1382
🟥 Series2	26.8066	27.519	28.3244	29.1479	29.7901
🟩 Series3	26.8532	27.5542	28.3119	29.0398	29.3568

Fig. 5 Lifting results PSNR versus decomposition level



	1	2	3	4	5
🟦 PSNR-Lena Image	31.4469	31.8372	32.3183	32.949	33.7251
🟥 PSNR-Urban Image	31.48	31.3788	32.3771	33.0415	33.7931
🟩 PSNR-Rural Image	31.5157	31.9257	32.4377	33.0679	33.617

Fig. 6 Adaptive lifting results PSNR versus decomposition level

5 Conclusions

The SPIHT compression scheme is not suitable for the compression of satellite imageries as the PSNR values are low and the quality of the reconstructed images is poor.The SPIHT scheme indicates that the PSNR values dramatically improves with decomposition levels and shows maximum PSNR for the 5 decomposition level. The classical lifting scheme shows dramatic improvement in the PSNR values derived for satellite imageries and indicated very high quality for the reconstructed images. The adaptive lifting schemes shows further improvement in PSNR values for satellite imageries compared to classical lifting schemes suggesting that they are best suited for compression of satellite imageries. For both the Lifting schemes the PSNR values does not show any significant improvement with decomposition levels and indicates that it is independent of decomposition levels.

References

1. Sweldens W (1997) The lifting scheme: a construction of second- generation wavelets. SIAM J Math Anal 29(2):511–546
2. Sweldens W (1995) The lifting scheme: a new philosophy in biorthogonal wavelet constructions. Proc SPIE 2569:68–79
3. Sweldens W (1996) The lifting scheme: a custom-design construction of biorthogonal wave lets. Appl Comput Harmon Anal 3:186–200
4. Cohen A, Daubechies I, Feauveau JC (1992) Biorthogonal bases of compactly supported wavelets. Commun Pure Appl Math 45:485–560
5. Said A, Pearlman W (1996) A new, fast, and efficient image codec based on set partitioning in hierarchal trees. IEEE Trans Circuits syst video Technol 6:243–250
6. Mohamed Ismail M, Baskaran K (2010) Clustering based adaptive image compression scheme using particle swarm optimization technique. Int J Eng Sci Technol 2(10):5114–5119
7. Hafiz DA, Sheta WM, Bayoumi S, Youssef BAB (2011) A new approach for 3D range image segmentation using gradient method. J Comput Sci 7(4):475–487

A Gene Expression Based Quality of Service Aware Routing Protocol for Mobile Ad Hoc Networks

Yeshavanta Kubusada, Giridhar Mohan, Kiran Manjappa and G. Ram Mohana Reddy

Abstract Mobile Ad Hoc Network (MANET) is a collection of infrastructure less multi-hop wireless mobile nodes which communicate together to achieve the global task. Despite lack of centralized control these mobile nodes still coordinate together to deliver the message to the destination node. MANET is gaining its popularity due to its easy deployment and self-organizing ability. In spite of its unique characteristics, mobility of mobile nodes causes frequent link breakups in MANET and thus makes route setup and maintenance a critical and challenging task. As real time and multimedia applications are increasing, there is a need of an efficient Quality of Service (QoS) aware routing protocol for MANET to support such applications. In the present work, the authors proposed an efficient QoS aware routing protocol for MANET based on upcoming Gene Expression Programming. In the proposed work, the information regarding the availability of resources is managed by a resource management module, which assists in selecting the resource rich path. Further, a theoretical proof is given for the proposed model for its correctness. The results are compared with the state of art artificial neural network and support vector regression methods from the performance evaluation point of view and the results are encouraging.

Keywords Gene expression programming · Mobile ad hoc networks · Sensitivity analysis · Artificial neural networks · Support vector regression

Y. Kubusada (✉) · G. Mohan · K. Manjappa · G. Ram Mohana Reddy
IT Department, National Institute of Technology Karnataka,
Surathkal, Mangalore, India
e-mail: yeshavanta.kp@gmail.com

G. Mohan
e-mail: giridharb54@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_28,

1 Introduction

Quality of service (QoS) provisioning can be done over different layers in the protocol stack, starting from the physical layer up to application layer. Each layer takes care of different measurements in QoS provisioning. For example, the physical layer takes care of transmission quality, the link layer handles the variable bit error rate, Network layer deals with the change in bandwidth and delay, transport layer focuses on the delay and packet loss due to transmissions errors and application layer aims at frequent disconnections and reconnections [1].

The Medium Access Control (MAC) protocol determines which node should transmit next on the broadcast channel when several nodes are competing for transmission on that channel. The existing MAC protocols for ad hoc wireless networks use channel sensing and random back-off schemes, making them suitable for best-effort data traffic. MAC layer aims at providing QoS guarantee for real-time traffic support. QoS implemented in the network layer aims at finding a route which provides the required quality. The metrics used to select the route are not only the number of hops along the route but also other metrics like delay, bandwidth, network life time and data rate.

QoS routing protocols search for routes with sufficient resources in order to satisfy the QoS requirements of a flow. The information regarding the availability of resources is managed by a resource management module, which assists the QoS routing protocol in its search for QoS feasible paths. The QoS routing protocol should find the path that consumes minimum resources while providing maximum possible throughput [1]. This determining of path that consumes minimum resources is done by gene expression programming.

2 Gene Expression Programming

Gene Expression Programming (GEP) aims to achieve an optimum solution, in the form of a mathematical expression, which relates the input and output variables of the data set. The GEP process begins with a random set of chromosomes that are produced to form the first generation and represents a solution to the problem in hand. Subsequent populations via genetic modification are descendants of this initial population. In order to form an initial population we have to choose the symbols for the chromosomes, that is, we must choose a set of terminals (the input variables) and a set of functions that are appropriate to solve the problem at hand. The resources used for QoS routing protocol here constitute the terminal set and we have confined operator set to Arithmetic operators: addition (+), subtraction (−), division (/), multiplication (*), sine, cosine, tangent, exponential function and random numerical constants. Next, we must provide a fitness function against which the fitness of each one is evaluated. Once the fitness values have been assigned to all the chromosomes of the population, then next task is to pick out the fittest chromosomes of the population to reproduce with modification. We have chosen selection mechanism based on roulette-wheel sampling in order to form the next population; the selected chromosomes are modified using the following genetic operators for our experiment.

- Mutation
- Inversion
- Transposition (Insertion Sequence, Root Insertion sequence, gene)
- Crossover (one-point and two-point)

Working of all the genetic operators has been explained in detail in [2]. The process of GEP has been schematically represented in Fig. 1. In this paper, we have used GEP technique to generate mathematical equations which predict possible throughput of the route. GEP is very efficient compared to other Evolutionary algorithms because it combines the advantages of both Genetic Programming (GP) and Genetic Algorithms (GA) [2] and it provides better results to the benchmark problems than both GP and GA [2]. Hence we have employed GEP for implementation.

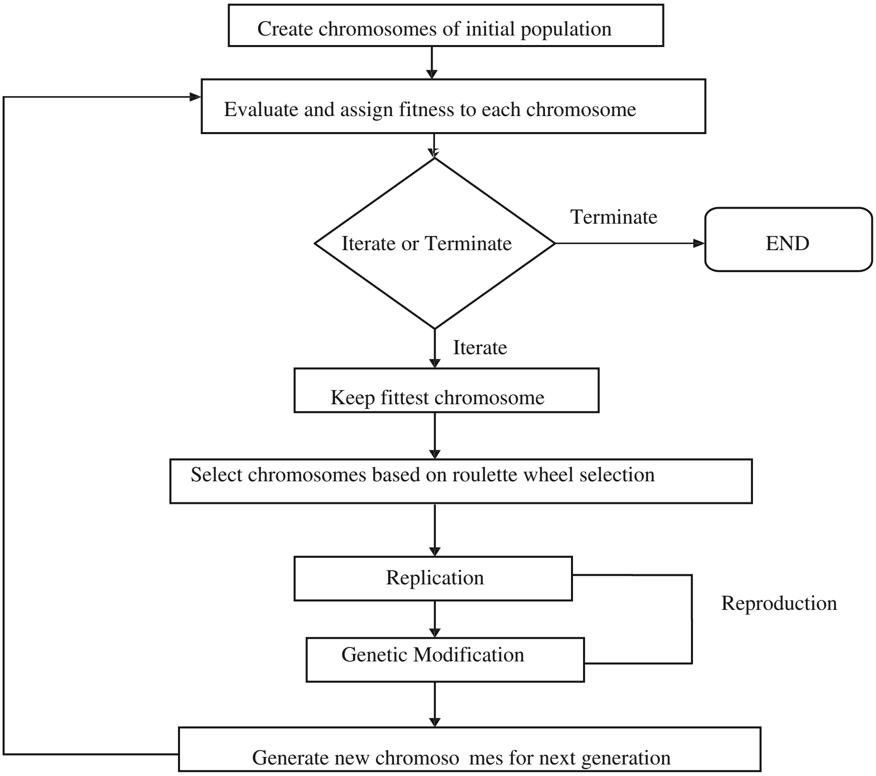

Create chromosomes of initial population
Evaluate and assign fitness to each chromosome
Iterate or Terminate
Terminate
END
Iterate
Keep fittest chromosome
Select chromosomes based on roulette wheel selection
Replication
Genetic Modification
Reproduction
Generate new chromoso mes for next generation

Fig. 1 Gene expression programming process

3 Methodology and Implementation

We have used GEP to obtain mathematical expression which predicts the possible throughput for the routes. Then we use this information to select the most feasible route to pass the packets.

3.1 Implementation Details

We have used the number of hops, Jitter and End-to-end delay as the three parameters in the dataset for a route to predict its output. Throughput is the parameter used as performance metric in varying scenarios to evaluate the performance of the proposed model. It is defined as the ratio of the number of the data packets received by the destination to those sent by the source. In this paper we predict the possible throughput of the path by using the mathematical expression generated from gene expression programming technique. This mathematical expression is obtained by considering Number of hops, end-to-end delay and jitter as parameters. A data set with number of hops, Jitter and end-to-end delay as input parameters and throughput as output is generated using network simulator 2. To prepare the dataset we first set up the network topology, then protocol to be used is determined and we have used transmission control protocol (TCP), then the type of application to be used is determined for which we chose FTP and finally we have used distance vector routing algorithm as routing protocol. Some of the important parameters set in the network simulator while generating the data set [1] are tabulated in the Table 1.
In simulation IEEE 802.11 DCF is used as MAC layer protocol. The data rate of wireless channel is fixed as 2 Mbps. Now we used gene expression programming technique to produce mathematical expression which given the values of end-to-end delay, number of hops and Jitter of the route could predict the possible throughput of the route. The data set had 80 samples of which we took 60 samples for training and all 80 samples for testing.

$$\mathrm{E_i} = (1/n)\sum_{j=1}^{n}\left(\left(P_{ij} - T_j\right) \div T_j\right)^2 \tag{1}$$

Table 1 Parameters used in network simulator 2

Parameters	Values
Simulation area	1000*1000 m
Number of nodes	100
Transmission range	250 m
Data payload	1500 bytes
FTP flows	1
Mobility	2 m/s

As we were employing gene expression technique we had to select a fitness function. Now since the output was throughput and had continuous values this problem was treated as function approximation rather than classification. Hence, we used the most preferred fitness function for function approximation [2] that is relative mean squared error. The fitness function is shown in Eq. 1 where P_{ij} is the value predicted by the individual chromosome i for fitness case j (out of n fitness cases) and T_j is the target value for fitness case j. Then sensitivity analysis has been carried out to find out which is the most sensitive input variable. Then based on the results of sensitivity analysis we even removed one input variable to get new mathematical expression which proved to be as efficient as the expression before sensitivity analysis. Then we compared the results obtained by GEP with benchmark algorithms such as artificial neural networks and Support vector regression.

3.2 *Sensitivity Analysis*

Sensitivity Analysis (SA) is the study of how the changes in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input [3]. It was carried out for each of the arithmetic expressions obtained in the previous section. The techniques used for SA are varied. But the technique that we have used is Parametric Sensitivity Analysis which is described in detail in [4].

4 Experiment and Results

The data set generated has three input variables namely Number of hops, Jitter, End-to-end delay which are denoted as d_0, d_1, d_2 correspondingly in the experiment and also while tabulating the results. The GEP technique has a number of parameters such as gene length, number of genes in a chromosome, number of chromosomes in a population of solutions and others, these values are tabulated in Table 2.

Table 2 GEP parameters

Parameter	Values
Exclusion level	1.1
Mutation rate	0.044
Inversion rate	0.1
Number of generations	5000
Population size	30
Linking function	+
Head length	9
Number of genes	3

We have also used constants in the chromosomes which helps in getting more accurate equations, which in turn increases the accuracy in predicting the throughput. By considering the output of sensitivity analysis as tabulated in Table 3 we eliminated the least sensitive input variable that is Jitter (d_2) and repeated the process of GEP. Hence, we obtained the final results. The final equation obtained is represented as three equations namely Eqs. 2, 3 and 4 and the constants used are tabulated in Table 4. The Equation obtained after the sensitivity analysis is as follows

$$\tan\left((G_1C_0)\,/\,(G_1C_1)\right) \times (d_0)/\sin G_1C_0/d_1 \tag{2}$$

$$\left((\tan\left(d_1 \times G_2C_0\right)) + \left(e^{G_1C_1}\right)\right) \times \left((d_1 - G_2C_0)\,/\,(d_1)^2\right) \tag{3}$$

$$(G_2C_0) - ((\tan\left((G_2C_1/d_1) - (d_1)\right)) - (\exp\left(\sin\left(d_0\right)\right))) \tag{4}$$

We obtain the throughput by adding the values from Eqs. 2, 3 and 4. We have used two constants in each gene; hence, there are totally 6 constants which are tabulated in Table 4.

Here the constants can be understood as follows G_1C_0 means first Gene and first constant and similarly G_2C_1 means second Gene second constant.

5 Comparison with Artificial Neural Networks and Support Vector Regression Methods

Artificial Neural Networks (ANN) are simplified models based on the biological learning process of the human brain [5]. The procedure most commonly used to train an ANN is a method known as back-propagation [6]. For training, a mean square error (MSE) of 10^{-5} and a maximum iteration number (epoch) of 50000 were used. The output function used is the sigmoid function. The structure of the ANN giving best results was 3-3-1 where the 3 represent the numbers of nodes in the input layer, the hidden layer had 3 nodes and the Output layer had 1 node. The initial weights and biases of the network were selected randomly. The results are shown in Fig. 2. One of

Table 3 Results of sensitivity analysis

Number of hops (d_0)	End to end delay (d_0)	Jitter (d_2)
70.2303535867	0.390253776908	0.0118885479

Table 4 Constants used in mathematical expressions

G_1C_0	G_1C_1	G_2C_0	G_2C_1	G_3C_0	G_3C_1
−0.633179	5.050354	−8.571106	5.050354	7.575378	7.987214

the main characteristics of Support Vector Regression (SVR) method is that instead of minimizing the observed training error, SVR attempts to minimize the generalized error bound so as to achieve generalized performance. This generalization error bound is the combination of the training error and a regularization term that controls the complexity of the hypothesis space. We have used two kinds of SVR namely nu-SVR, and epsilon SVR .The results are shown in Fig. 2 and we can see that only GEP and ANN closely follow the actual values of the throughput but GEP has a clear advantage because it provides a Mathematical expression.
x-axis represents the number of data samples and y-axis represents the throughput in kilobytes per second. And then Series 1 represents target value or the actual experimental value of throughput, Series 2 the values predicted by GEP, Series 3 the values predicted by ANN, Series 4 the values predicted by Epsilon SVR and Series 5 the values predicted by nu-SVR. In Fig. 2 we can see that GEP predicted values closely follow the actual values and ANN is also pretty accurate, but nu-SVR and epsilon SVR do not show such good results. But still we can argue that GEP is better than ANN since it gives us mathematical equations unlike ANN.

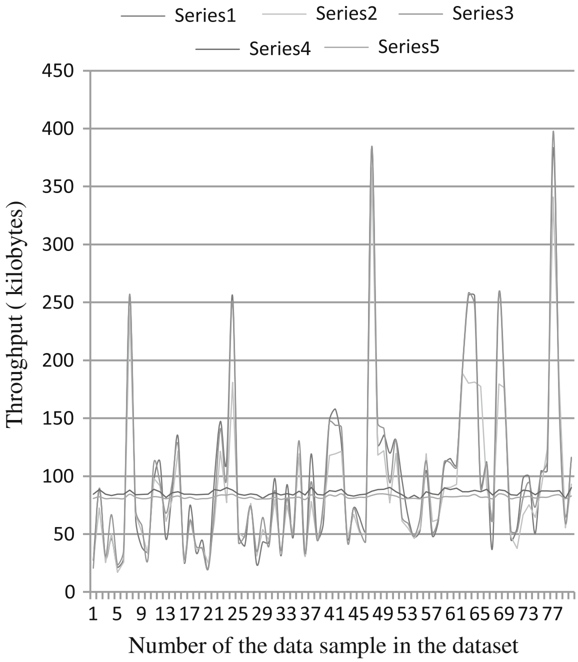


Fig. 2 Approximation graph showing the values predicted by GEP, ANN and SVR

6 Conclusion

Many applications have stringent QoS requirements such as throughput, end-to-end delay, jitter and network lifetime and other. However existing QoS routing solutions deal with only or two of the required QoS parameters. To overcome this problem, new protocols that satisfy the QoS parameters are needed. Gene Expression Programming which is a new nature inspired algorithm, can be implemented in diverse field of networks such as MANETs. The key to the implementation is the flexibility and simplicity that GEP offers. The results have demonstrated that the performance of GEP is same as that of ANN and better than SVR. However, ANN needs normalized dataset whereas GEP method utilizes the original data set without normalization.

References

1. Mohapatra P, Jian L, Gui C (2003) QOS in mobile and adhoc networks. IEEE Wirel Commun Mag 10(3):44–52
2. Ferriera C (2006) Gene expression programming, mathematical modelling by artificial intelligence, 2nd edn. Springer.
3. Satelli A (2002) Sensitivity analysis for importance assessment. Risk Anal 22(3):579–590
4. Cho K-H, Shin S-Y, Kolch W, Wolkenhauer O (2006) Experimental design in systems biology based on parameter sensitivity analysis with Monte Carlo simulation: a case study for the TNF α mediated NF- κB transduction pathway. Brief Bioinform 7:364–374
5. Anderson JA (1972) A simple neural network generating an interactive memory. Math Biosci 14:197–200
6. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by backpropagating errors. Nature 323:533–536

User Behavior and Capability Based Access Control Model and Architecture

Meriem Zerkouk, Abdallah Mhamed and Belhadri Messabih

Abstract Owing to ambient intelligence and context awareness in nowadays environments, the provided services become more pervasive and personalized according to the user's profile. With the growing healthcare and wellbeing context aware applications, modeling security policies become an important issue in the design of future access control models. This requires rich semantics using ontology modeling for the management of services provided to dependant people. However, current access control models remain unsuitable due to lack of completeness, flexibility and adaptability to the user capability and behavior. In this paper, we propose a novel adaptable access control model (UBC-ACM) and its related architecture (UBC-ACA) in which the security policy is based on the user's behavior and capability to grant a service using any assistive device within intelligent environment. The design of our model is an ontology-learning and evolving security policy for predicting the future actions of dependent people. This is reached by analyzing historical data, contextual data, and user behavior according to the access rules that are used in the inference engine to provide the right service according to the needs of users.

1 Introduction

Pervasive systems provide an assistive environment allowing dependent people to perform their required services in their living spaces for various applications (healthcare, smart home and transport). It fits to the needs of disabled and older people by

M. Zerkouk (✉) · B. Messabih
University of Sciences and Technology, Oran, Algeria
e-mail: zerkouk.meriem@gmail.com

B. Messabih
e-mail: abdallah.mhamed@it-sudparis.eu

A. Mhamed
Institut Mines-Télécom / Télécom SudParis, CNRS Samovar UMR, 5157Evry, France
e-mail: bmessabih@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_29,

making their life easier to overcome barriers for studying, working and living activities. The adaptability must rely on services which become more and more personalized with respect to the assistive devices technology. Development and innovation of new assistive technologies and services are in continuous progress to provide help for the daily life activities. Even if provided services are more and more personalized according to the profile and preference of users, they still do not take the user capabilities into account. The access control is handled by the development of context aware based access control models for pervasive systems. The security policy must be adaptive to the potential changes which may occur over the space and the time. This is accomplished by extending the most popular Role based access control model (RBAC) [1] and Organization Based Access Control model (OrBAC) [7]. The permissions are assigned according to the validity of context, which is a key element in the design of access control models taking into account the growing situation in the environment. According to our literature review the current access control policies do not take into account the user impairments nor the behavior of users. Our work is motivated by the following challenging tasks:

1°) Providing a better identification of users based on their capabilities and behavior to ensure more suitable personalization.

2°) Using the different gathered and inferred data by the security policy specification to ensure security services.

3°) Assigning correctly the users having similar characteristics and deduce the suitable decision about the user.

In this paper, we present and describe a novel adaptive model. The model is based on ontology learning, enriching and evolution which support continuous learning of behavior and capability patterns. We designed an initial ontology driven knowledge base representing the access rules, context, behavior pattern, services, devices and environment. We used ontology modeling to ensure sharing, reuse, interoperability, flexibility, adaptability of the security policy.

The reminder of this paper is organized as follows: Section 2 analyzes the requirements of access control models and their limitations in pervasive systems. Section 3 details our proposed ontological model while Sect. 4 gives the description of the related architecture. Section 5 illustrates the implementation of our approach through a scenario and the last section concludes the paper.

2 Related Work

From our literature study, RBAC seems as a standard and reference for the design of access control model in pervasive environments. The model is based on a set of users, roles, permissions, operations, object entities, user-role and role-permission relationships [1]. The model is defined by four components: core RBAC, hierarchical RBAC, static separation of duty and dynamic separation of duty. Context is a key challenge in ubiquitous computing. Therefore, the most popular definition of context extracted from [2] is: *Context is any information that can be used to characterize the* *situation of an entity. An entity is a person, a place or an object that is considered as relevant to the interaction between a user and an application themselves.* Furthermore, the contextual data varies according to the context awareness environment like hospital, home, and work place. Context aware based Access control rely on context data to assign the permission to the users (roles) in the right situation which makes the model dynamic according to the change of context over the time.

Extended RBAC models [3] are based on context awareness. Their aim is to improve RBAC by assigning the right access more dynamically. The access is based on the validity of the context by adding to RBAC a single contextual data which is spatial, temporal or environmental [4–6].

The OrBAC model is designed to overcome existing problems in extended RBAC models. OrBAC [7] adds an organizational dimension and separates between the concrete level (user, object, action) and the abstract level (roles, views, activities). It also models alternative policies (permission, prohibition and obligation, recommendation). This model incorporates different context data which can be historic, spatial, temporal or declared by user. The weakness of the model is the lack of handling the interoperability and the heterogeneity. Multi-OrBAC [8] is an extension of the OrBAC model designed for multi-organizational, its drawback lies in the fact that each organization must know the resources of the other. Poly-OrBAC [9] addresses this problem by integrating the OrBAC model to represent the internal policies of each organization and web services technology to ensure interoperability between organizations.

3 Proposed Model

In order to implement the security services, we provide an intelligent security policy specification framework following the main four steps:

Constructing user behavior profile model: this step is necessary to assign behavior classes to users using a classifier process. The classifiers are based on the historic data to build the behavior models.

Modeling: it consists to represent data on standard format by using ontologies to ensure the interoperability, the sharing and the reuse of security policy.

Reasoning: it uses the current captured data and the inference rules stored in the database to deduce a new knowledge and to check the consistency of the ontology.

Evolving: it consists to learn the data provided over the time from different sources in order to update the behavior classes.

3.1 Security Policy Modeling

In this paper, we propose a new User Behavior and Capability based Access Control Model (UBC-ACM) which provides an adaptable access control in smart environment where the users have specific behavior, profile and capability. Therefore, the assistance is required to adapt services according to the users using assistive devices. For this, we classified users according to behavior, profile, current context situation, services, devices and environment. In order to take the real entities, we use the semantic web technologies, the ontology is described with Web language (OWL) and Semantic Web Rule language (SWRL) and the Simple Protocol and RDF (Resource Description Framework) Query Language (SPARQL) queries to ask and access to the data, these tools are used to define, represent and implement our proposed model in smart environment.

As illustrated in Fig. 1 the ontology is built using four principal entities: security policy, user, service, device and environment.

Our (UBC-ACM) model is defined as an ontology that expresses the security policies in such smart environment. In order to ensure an adaptive security policy, we need to take into consideration five main classes: user, device, service, environment and security policy.

The Policy class permits to ensure the authentication, the access control, the trust, the priority and the conditions. Authentication subclass its goal is to identify correctly the user using behavior subclass and authentication credentials (username/password, bag or biometric data). The behavior identification of the user is needed to deduce information about the trust value while the capability identification is used to assign a priority value for user. The access control subclass defines a set of authorization (permit or deny), prohibition, recommendation and obligation which the decision is assigned according to the behavior, profile and context of the user. The policies are specified as a set of rules using SWRL form.

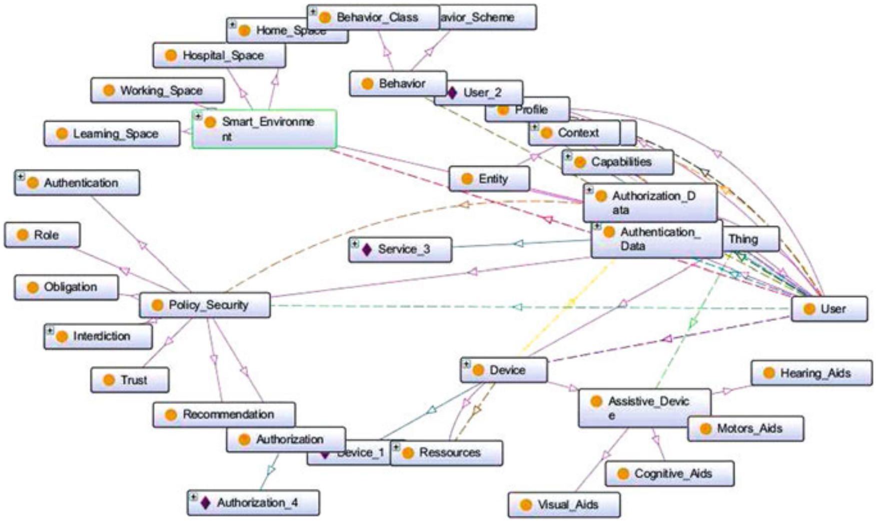


Fig. 1 Ontology model (UBC-ACM)

The User class: This entity aims to describe any person living with a disability due to any impairment, illness or simply aging.

The Profile subclass is used to provide fined grain/valuable data which help to distinguish one user from other by using both static profile (personal data, capabilities, hobbies) and dynamic profile which include (interests, preferences, opinions, moods). The capabilities subclass distinguishes cognitive, visual, hearing and motor impairments.

The Behavior subclass after having identified the specific profile and the capability of the person then we should check the behavior with respect to the background while classifying them into behavior class to recognize the right people.

The Context subclass describes the current state of the user in the environment that includes activity describing the current task, location characterizing the place of the user and time describing years, months, weeks, days, hours and minutes.

The Service class describes an entity solicited by an appropriate user capability and behavior.

The Device class aims to describe any item or piece of equipment used to maintain or improve the functional capabilities of a person with a disability. It focuses on assistive devices subclass including cognitive aids, hearing aids, visual aids, motor aids and resources subclass including audio, video and text.

The Environment class describes the intelligent and adaptable spaces that support the assistive services and devices for appropriate user profile capabilities like smart home, work space, learning and healthcare. This class specifies the indoor and outdoor subclasses.

3.2 Reasoning

In order to infer the suitable authentication and access control decisions, we use a semantic reasoner that is able to infer the decision from a set of asserted observations about the dependant person located in smart environment. Therefore, these persons needs to access to smart services that take into account their impairments by using assistive devices. We specify semantic rules by the means of semantic web technologies. Usually, the rules respect the predicate logic and are specified in the form <if conditions then conclusion> to perform the reasoning. Semantic reasoner aims to check the consistency of our proposed ontology model and derive high and implicit knowledge about the situation, the profile, the capability, the authentication and the access control of the user.

We choose to implement the reasoning by using a complete open source OWL-DL reasoner as pellet, racer that conforms to our ontology modeling.

3.3 Rules specification

Our ontology is rich by using the contextual data, the capability and the behavior from users in such intelligent environment. This process is powered by the heterogeneous data captured from different sensors. We focus on ontology developed and rule base to perform the reasoning in order to deduce a new implicit knowledge which the rules are expressed as SWRL form. In this section, we organize the rules on contextual, profile capability, behavioral, authentication and access control rules specification. Contextual rules permit to provide a contextual knowledge about the user situation, time, location, device and environment. These data will be used in the remainder rules specification to deduce other high implicit knowledge. Profile capability rules, when a user asks for a particular service and device, we can deduce the user capability. For example, if the user uses an assistive device like hearing aids then we conclude that the user has hearing impairments and needs to input/output video resources. Behavioral rules provide to conclude other data about the profile, the authentication and the access control, if the user has a recognized behavior class. Authentication rules permit to deduce the decision according to the recognized user behavior and to confirm the decision, we use the traditional authentication means. Authorization rules need to validate the inferred data about the behavior, capability, context, service and device to assign the suitable decision to users.

3.4 Query specification

Once, the data are modeled and the rules are specified with rich semantics, therefore, we may query the stored data using a semantic query language like SPARQL. In our approach, we use two kinds of queries. The first one to specify an authentication query that requires a user name, password, behavior model to pass the query to decision making level based on inference engine. The second to send an authorization query to inference engine with the specification of the situation, profile, capability and behavior.

4 Proposed Architecture

In this section, we propose a new User Behavior and Capability based Access Control Architecture (UBC-ACA). Our model is based on semi-automatic management architecture as illustrated in Fig. 2. It requires in input the profile, behavior model and current state of the user and provides on output an authentication and authorization decision access.

The architecture is built on three levels.

Acquisition layer: it captures the contextual data from hardware, software and combination of sensors on raw format that will be passed to the upper layer through an interface.

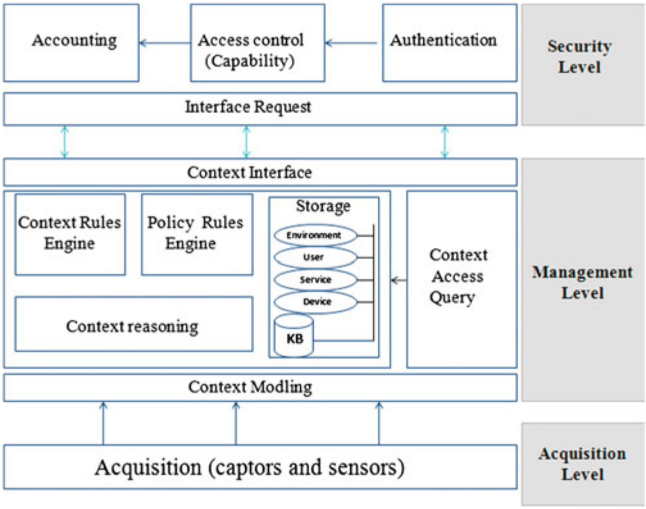


Fig. 2 Proposed architecture (UBC-ACA)

Management layer: it is the core of architecture that provides modeling, reasoning storage and querying process.

Security layer: It is based on AAA (Authentication, Authorization and Accounting) architecture to ensure adaptive security policies in smart environment.

Authentication: to ensure an adaptive policy, the correct identification is required by checking the capability, the behavior then using the authentication means. The improved identification provides better personalization.

Authorization: once the user is identified correctly according to his profile and class of behavior, capability and the current contextual data then the suitable decision will be derived about the user.

Accounting: Tracking the user activities to detect the incorrect activity because the dependent users require a high assistance in their smart home to update the historic file.

5 Scenarios

Using our access control system, we go through three phases to derive the right decisions: modeling, reasoning and asking the ontology model.
Scenario 1: Deaf Person
Situation: person having hearing impairment. So, it needs to specific visual devices to perform their tasks as reading and daily activities….
Action: Visual alert.

Scenario 2: Blind Person
Situation: person having visual impairment. So, it needs to specific audible devices to perform their tasks as reading and daily activities easily ….
Action: Audible alert.
Scenario 3: Alzheimer Person
Situation: It is a progressive disease which the person destroys cognitive abilities and it suffers about memory loss, abnormal behavior, a change in personality and an increase of anxiety.
Action: signaling the emergency situation and correcting the future activity.

5.1 Modeling the Security Policy Ontology

We propose an ontology model taking into account the three case studies: deaf person, blind person, Alzheimer person. These persons live in smart environment; they ask different services, use different devices according to the disability type with different context situation.

5.2 Reasoning

We study the different use cases and showing the importance of an adaptive access control and the role of tracking the behavior, the profile capability in such adaptive access control system.

5.3 Access Control Policies

The access control is ensured according to the assignment of the users to behavior and capability groups then we check the valide time, location, device, service and environment to assign the “permit” or “deny” decision.

We use this rule to assign the user into groups according to their behavior and capabilities.

```
HasRecognizedbehavior(?u, class1) ,HasCapability(?u,
“hearing”)->BehaviorCapability(?u,Group1)
HasRecognizedbehavior(?u, class2) ,HasCapability( ?u,
“visual”)->BehaviorCapability(?u,Group2)
HasRecognizedbehavior(?u, “class2”) ,HasCapability( ?u,
“cognitive”)->BehaviorCapability(?u,Group3)
BehaviorCapability(?u,Group2)
,AskedService(?Group2,?S),UsedDevice(?Group,?D),
HasContext(?C)->has Access (?u, permit).
```

Then we send an access control query to collect the decisions after the reasoning process.

Following this scenario, we have been shown the importance of taking into account the user behavior and capability when specifying the security policy and how adapting both authentication and access control security services from the needs of users.

6 Conclusion

The adaptation to the user needs becomes the key issue for providing the access to personalized services. In this paper, we have proposed an access control model (UBC-ACM) based on behavior and capability. The model supported by our proposed architecture (UBC-ACA) is involving three design layers (acquisition, management and security).

Our security policy was developed with an ontology which expresses the concepts and rules security policy. The selected scenarios illustrate how the visual, the hearing and the cognitive capabilities can affect the design of security policy to ensure both authentication and access control security services. Motivated by these scenarios, we are planning to deploy our model and architecture in real living area of physically impaired people.

References

1. Ferraiolo D, Sandhu RS, Gavrila S, Kuhn D, Chan-dramouli R (2001) A proposed standard for role-based access control. ACM Trans Inf Syst Secur 4(3):224–274
2. Dey A (2001) Understanding and using context. Pers Ubiquit Comput 5(1):4–7
3. Asmidar A, Roslan I, Jamilin J (2009) A review on extended role based access control (E-RBAC) model in pervasive computing environment. International conference on intelligent pervasive computing, IEEE, pp 533–535
4. Emami SS, Amini M, Zokaei S (2007) A context-aware access control model for pervasive computing environments. In: International conference on intelligent pervasive computing, IEEE, 51–56
5. Yao H, Hu H, Huang B, Li R (2005) Dynamic role and context based access control for grid applications. In: Proceedings of the sixth international conference on parallel and distributed computing, applications and technology (PDCAT'05), December 2005, IEEE, pp 404–406
6. Filho JB, Martin H (2009) A generalized context-based access control model for pervasive environments. In: Springle in '09, pp 12–21
7. Miège A (2005) Definition of a formal framework for specifying security policies: the Or-BAC model and extensions. Ph.D, Computer security, ENST-INFRES computers and networks, ENST
8. Abou El Kalam A, Deswarte Y (2006) Multi-OrBAC: a new access control model for distributed, heterogeneous systems andcollaborative. IEEE symposium on systems and information security, Sao Paulo, Brazil
9. Abou El Kalam A, Deswarte Y, Baïna Kaaniche A (2009) PolyOrBAC: a framework for critical infrastructure security. Report LAAS No. 09087, March 2009, p 28. (to appear in International Journal on Critical Infrastructure Protection, Elsevier)

Congestion Adaptive Most Favorable Control Routing in Ad Hoc Networks

S. Subburam and P. Sheik Abdul Khader

Abstract Routing protocols for mobile ad hoc networks have been researched extensively in recent years. However, most routing protocols attempt to discover a new route when congestion is said to have occurred. Such an approach is suggested to cause further delay, high packet loss and requires significant overhead for identifying a new path. An ideal routing protocol is expected to detect congestion as well as be adaptive to network congestion. Adaptability to congestion helps to increase both effectiveness and efficiency of routing. This paper proposes an improved AODV routing protocol that attempts to discover a new path if congestion is likely in a present path, using a routing algorithm called CAMFCR (congestion adaptive most favorable control routing). Initially, CAMFCR segregates the network in to five regions according to the means of neighbors. After segregation of the network, it initiates a most favorable route discovery process. Once new paths were identified, all the nodes in primary path periodically calculate its queue_status at the node level. In this approach, each node will send a warning message to its previous node if the former node is likely to be congested. In response, the latter node will attempt to identify a new path for the approaching packets to be sent through, by applying a most favorable route discovery method. Thus, CAMFCR improves the network performance in terms of reducing delay and routing overhead and increasing packet delivery ratio. The analysis using Ns-2 simulator revealed that CAMFCR is superior over AODV and CAAODV routing schemes.

S. Subburam (✉)
Satyabama University, Chennai, Tamil Nadu, India
e-mail: subburam@bsauniv.ac.in

P. Sheik Abdul Khader
BSA University, Chennai, Tamil Nadu, India
e-mail: hodca@bsauniv.ac.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_30,

1 Introduction

Mobile ad hoc network (MANET) is an autonomous network of mobile nodes that communicate with each other over a wireless path. In this network mobile node acts as a host when requesting/providing information from/to other nodes in the network, as well as acts as a router when attempting to discover or maintain paths for other nodes in the network. Each MANET node performs the routing function independently. Usually, routing protocols are based on a simplistic form of broadcasting called flooding where every node in the network retransmits every received packet exactly once; however, it may lead to a serious problem often called the broadcast storm problem [1–3]. Usually in AODV routing protocols, the routing table is checked to see whether the route is available so that the source node can transmit a packet to the destination node. If a route was found, then the node proceeds to transmit the packet; else it broadcasts a RREQ (route request) packet to the destination node, which on receipt of it sends a RREP (route reply) to the source node. One of the fundamental functions of an ad hoc network is congestion control. The objective of congestion control is to limit the delay and buffer overflow caused by network congestion and provide better performance in the network [4].

2 Congestion Adaptive most Favorable Control Routing

CAMFCR is a unicast routing protocol for MANET. It reduces network congestion by reducing the unnecessary flooding of packets and finding a congestion-adaptive path between the source and the destination. In this paper, we propose the design and an in-depth evaluation of the CAMFCR protocol. Initially, the CAMFCR protocol segregates the network in to low sparse, medium sparse, medium, medium dense and high dense regions using mean values of neighbors. Afterwards, it initiates the most favorable route discovery process. Once a new path has been established, it proceeds to transmit data packets through the network. If any of the nodes is likely to be congested, CAMFCR utilizes the non-congested predecessor node and initiates the most favorable route discovery process.

Thus, the main objective of CAMFCR is to discover a congestion-free path between the source and the destination nodes, while reducing the overhead and avoiding flooding of packets. To understand the CAMFCR protocol requires recognizing the following aspects:

1. Most Favorable Route discovery and Network classification
2. Early congestion detection
3. Working principle of CAMFCR

2.1 Most Favorable Route Discovery and Network Classification

In a MANET, several system parameters affect network performance, such as node, node density and congestion, because the nodes in a MANET are randomly situated and the topology changes frequently. A node's neighbor information is the pointer to decide whether or not a current node is in dense area; the information is collected by broadcasting "hello" packet every second for only one hop. This packet will ensure that every node has an updated neighbor list. With respect to the number of neighbors for each node, the rebroadcast will be dynamically adjusted. From extensive simulation studies, three values of average number of neighbors are determined: the minimum average number of neighbors (Avgmin), the average number of neighbors (Avg) and the maximum average number of neighbors (Avgmax) [1, 2]. Equations (1–3) are used to calculate the three threshold values, which will be used to identify whether the region is in high dense, medium dense, medium, low sparse or medium sparse. Let N be the number of nodes in the network; N_i be the number of neighbors for node D_i.

To find the average number of neighbors:

$$\tilde{n} = \frac{\sum_{i=1}^{N} N_i}{N} \tag{1}$$

where ñ is the mean average number of neighbors; is the number of neighbors of each node; N is the total number of nodes.

To find the minimum number of neighbors:

$$\tilde{n}\ \text{min} = \frac{\sum_{i=1}^{r} N_i}{R} \tag{2}$$

where ñ is the mean minimum number of neighbors; is the number of neighbors of each node below average; R is the total number of nodes whose neighbors are below average.

To find the maximum number of neighbors:

$$\tilde{\text{n}}\ \text{max} = \frac{\sum_{i=1}^{k} N_i}{K} \tag{3}$$

where ñ is the mean maximum number of neighbors; is the number of neighbors of each node above average; K is the total number of nodes whose neighbors are above average.

In this phase, the source will transmit a packet to destination and find the path; it broadcasts the RREQ to the destination node. Once the packet is received at the destination it compares with the threshold and identifies the region of the node. if $(n < \tilde{n}\min)$ then the node lies in low sparse area, if $(n \geq \tilde{n}\min \;\&\&\; n < \tilde{n}\max)$ then the node lies in medium sparse area, if the node lies in $(n == \tilde{n}\max)$ then the node is in medium area, if the $(n > \tilde{n}\max \;\&\&\; n \leq \tilde{n}\min)$ then the node lies in medium dense area and if the node lies in $(n > \tilde{n}\min)$ then the node lies in dense area based on the regions the rebroadcast of the packet takes place else rebroadcast is stopped. From the Fig. 2, we are finding out the total number of possible rebroadcasts of an Route request packet (Total no of ReBroadCast) using the Most favorable control routing. We understood that Most favorable control routing broadcast less number RREQ compare with normal AODV algorithm.

From Fig. 1, we know that the total number of nodes is equal to 15. The total number of neighbors will be calculated based on one-hop neighbors of each node and we get the value of 47.

Therefore,

$$\begin{aligned}
\text{AVG} &= \text{Total number of neighbors/ Number of nodes}\\
&= 47/13 = 4\\
\text{MIN} &= \text{Total number of neighbors below average/Number of nodes below average}\\
&= 26/10 = 3\\
\text{MAX} &= \text{Total number of neighbors above average/Number of nodes above average}\\
&= 12/2 = 6
\end{aligned}$$

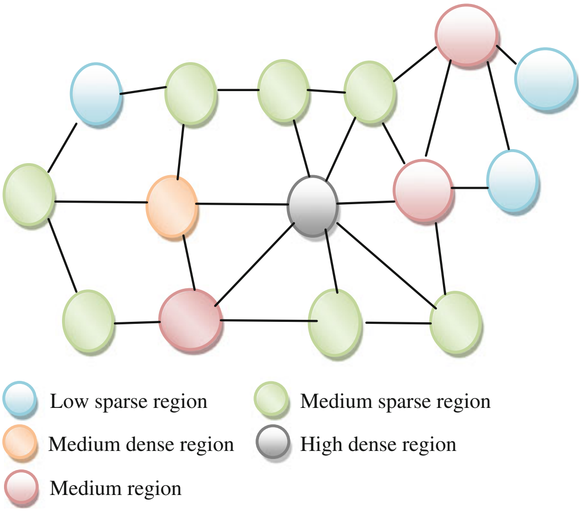


Fig. 1 Network classification

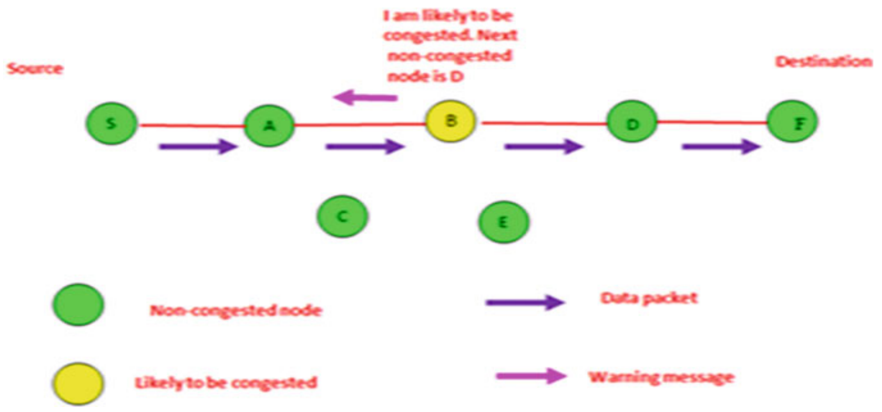


Fig. 2 Intermediate node buffer size below threshold

2.2 Early Congestion Detection

Congestion in a network may occur at any node when the number of packets coming to that node exceeds its buffer capacity. A variety of metrics could be used at a node to monitor congestion status [5]. For instance, one such metrics could be based on the percentage of packets discarded for lack of buffer space and the average queue length. Every second, a node checks the occupancy of its path-layer queue to detect the congestion well in advance. The early congestion detection technique proposed here is based on a queue management algorithm with an optimization of random early detection (RED) model that makes use of direct measurement of congestion status in a network [6].

Expressions (4–6) are used to set the minimum threshold and maximum threshold values for the queue length. The fixing of low threshold, medium threshold and high threshold values depends on the desired average queue size, and the fixing of threshold values will play a vital role in predicting congestion. In our algorithm, we chose to fix the low threshold at 40 %, the medium threshold at greater than 40 % and less than 80 %, and high threshold at greater than 80 %.

$$\text{LowTH} = 40\,\%\ \text{queue\_size} \tag{4}$$

$$\text{MediumTH} = \text{greater than } 40\,\% \text{ and less than } 80\,\% \tag{5}$$

$$\text{HighTH} = \text{greater than } 80\,\% \tag{6}$$

If the queue_size is less than LowTH, then the node can be classified to be in safe zone (green color), greater than LowTH but less than HighTH is classified as likely to be congested zone (yellow color), and greater than HighTH is classified as congested zone (red color).The above calculated value will be used to find out weather the particular node is non congested or likely to be congested or it is in congested zone.

Based on this value congestion status will be informed to previous non congested node.

2.3 Working Principle of CAMFCR

Our proposed protocol of CAMFCR attempts to prevent network congestion. The working principle of CAMFCR involves congestion monitoring, bypass discovery, traffic splitting and failure recovery. Briefly, every node present on a path notifies its previous node when prone to be congested. The previous node uses a "bypass" path for bypassing the potentially congested node. Traffic is split probabilistically between these two paths, primary and bypass, thus effectively lessening the chances of congestion occurrence.

2.3.1 Bypass Discovery

The predecessor node, on receipt of the update packet, will update its routing table. If that node is "likely to be congested" or "congested", the predecessor node will find a bypass path to the first non-congested node in the primary path. Figure 2 depicts the scenario where intermediate node B's buffer size is below the threshold. It sends a warning message to its predecessor node B. It also sends information of the next non-congested node (node D) in the primary path through the warning message.

2.3.2 Traffic Splitting

Once the predecessor node in Fig. 3 discovers the bypass path, it proceeds to split the traffic between the primary path and bypass path. Figure 4 depicts the scenario where the traffic is split based on the congestion status.

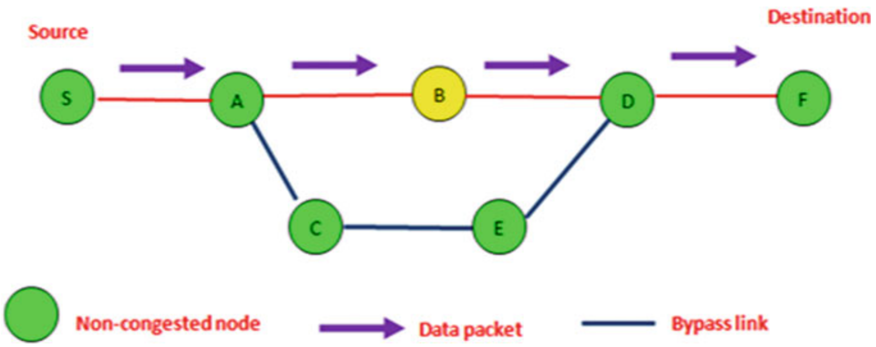


Fig. 3 Predecessor node finding the bypass path

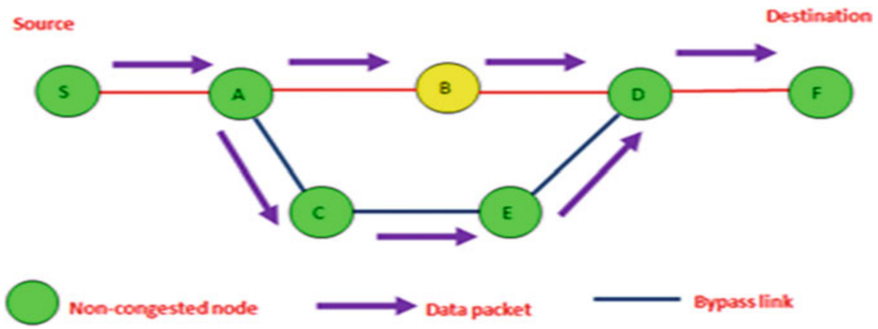


Fig. 4 Traffic splitting between the primary path and bypass path

Pseudo code for congestion status of the current node describe how the congestion will be detected before it occurs and explain about how it will be avoided in each and every node using the bypass path.

Pseudo code for congestion status of the current node

```
     If(congestion status of current node is  green)
        Current Node is non congested . So do not
         inform to its Non Congested Neighboring node
     Else
        Current node is congested because
        Status of Current node is either yellow  or red.
   If(status of current node is either yellow or red)
  Set the routing table of bypass value=1
  Send the routing table of bypass value and cogestion
  status of current node to the neighboring non
  congested node .Neigboring non congested node
  receive the details from congested node.
  Store this details into Alternate path routing table
If((string compare (rp->hop_status, “yellow”)==0)
   ||(string compare (rp-> hop-> status, “red”)==0))
       If(rt->bypass !=1)
        set bypassflag=1
     Call the route-down procedure
     Initiate route discovery process
     If(found)
              Bypass-path= True
              Primary-path =True
              Use bypass path and primary path for communication
    Else
              New-path= False
              Existing-path = True
              Use primary path for communication
```

3 Performance Metrics

For our experiments, we used a simulation model based on *Ns-2* [7]. The mobility model used the *random waypoint* model [8] in a rectangular field. Field configuration was 1400 × 1400 m with 100 nodes. Here, each node started its journey from a random location to a random destination at a randomly chosen speed (uniformly distributed between 0 and 20 m/s). The CBR sources were sending data packets to the destinations at fixed a rate of 16 packets per second. On reaching the destination, another random destination was targeted after a pause. We varied the pause time, which may have affected the relative speeds of the mobiles. Simulations were run for 300 simulated seconds for 100 nodes.

Packet delivery fraction—Figure 5 shows the ratio of data packets delivered to the destination to that generated by the CBR sources.

Average end-to-end delay—Figure 6 includes all possible delays caused by buffering during route discovery, queuing at the interface, retransmission delays at the MAC, propagation and transfer times.

Normalized outing load—Figure 7 shows the number of routing packets "transmitted" per data packet "delivered" at the destination. Each hop-wise transmission

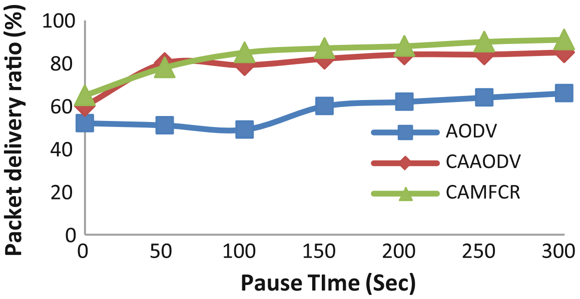


Fig. 5 Packet delivery ratio

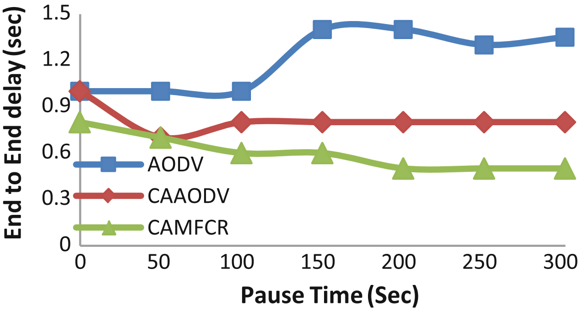


Fig. 6 End-to-end delay (s)

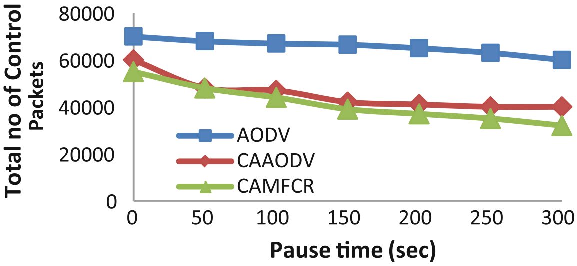


Fig. 7 Routing overhead

of a routing packet is counted as one transmission. The first two metrics are the most important metrics for best-effort traffic. The routing load metric evaluates the efficiency of the routing protocol.

4 Conclusion

In this paper, we proposed a novel way of accomplishing congestion control called CAMFCR in wireless multihop networks. This approach attempts to detect congestion before occurring in the first place rather than dealing with it reactively after occurring. The previous uncongested node, on request from a node that is likely to be congested, finds a bypass path to the destination immediately by applying a most favorable route discovery method, and the packets will be sent through that primary path and congestion is eliminated as a result. A key advantage in CAMFCR design is that it desists from usual flooding, which results in a very high overhead, especially in large dense networks. CAMFCR employs the most optimal control routing method for reducing the RREQ overhead during the route discovery operation. This approach can substantially reduce the overhead as compared to the existing flooding techniques. The outcomes of the proposed implementation were analyzed using NS2 network simulator. The performance of the proposed CAMFCR was compared against the existing AODV and CAAODV. The simulation results revealed that the proposed protocol gives better performance in terms of packet delivery ratio, end-to-end delay and overhead.

References

1. Yassein MB, Khalafand MB, Al-Dubai AY (2010) A new probabilistic broadcasting scheme for mobile ad hoc on-demand distance vector (AODV) routed networks. J Supercomput 53:196–211
2. Yassein MB, Nimer SF, Al-Dubai AY (2011) A new dynamic counter-based broadcasting scheme for mobile ad hoc networks. Simul Model Pract Theory 19:553–563
3. Ni S-Y, Tseng Y-C, Chen Y-S, Sheu J-P (1999) The broadcast storm problem in a mobile ad hoc network. In: Proceedings of ACM/IEEE Mobicom'99, pp 151–162

4. Floyd S, Jacobson V (1993) Random early detection gateways for congestion avoidance. IEEE/ACM Trans Netw 1:397–413
5. Senthilkumaran T, Sankaranarayanan V (2010) Early detection congestion and control routing in MANET. In: Proceedings of the seventh IEEE and IFIP international conference on wireless and optical communications networks (WOCN 2010), pp 1–5
6. Floyd S, Jacobson V (1993) Random early detection gateways for congestion avoidance. IEEE/ACM Trans Netw 1:397–413
7. Perkins C, Belding-Royer E, Das S (2001) Performance comparison of two on-demand routing protocols for ad hoc networks. IEEE Pers Commun 8:16–28
8. Broch J, Maltz DA, Johnson DB, Hu Y-C, Jetcheva J (1998) A performance comparison of multi-hop wireless ad hoc network routing protocols. In: Proceedings of the 4th international conference on mobile computing and networking (ACM MOBICOM'98), pp 85–97

An Improved Blind Channel Estimation Based on Subspace Approach for OFDM Systems Under Fast Time Varying Conditions

Zaier Aida and Ridha Bouallegue

Abstract In this paper, an improved channel estimation scheme is proposed for fast time varying channel in an OFDM context. This scheme is based on a subspace approach for blindly identifying the OFDM frequency selective channel and it will be shown from simulations that this approach is preferable than the other methods based on the training sequences which are expensive in term of channel efficiency and capacity. So, in this article and based on the simulations results, we will try to show the great effect of this approach when added in the system model in term of bit error rate BER and mean square error MSE versus the signal to noise ratio SNR.

Keywords OFDM systems · Time varying channel · Blind estimation · Subspace approach

1 Introduction

Orthogonal Frequency-Division Multiplexing (OFDM) systems had improved the channel require in term of bit rate and invoked many research to turn this system as efficient as possible. It was shown also from the continuous studies of this promoting system that it is also an effective technique that produces a high spectral efficiency and

Z. Aida (✉)
Department of Telecommunications, 6'Tel Research Unit, Innov'Com laboratory,
National Engineering School of Tunis, Tunisia, 6 Farhat Hached Road, 6010 Metouia, Tunisia
e-mail: zaieraida@yahoo.fr

R. Bouallegue
Sup'COM Tunisia 6'Tel Research Unit, Innov'Com laboratory,
Higher School of Communications of Tunis, Tunis, Tunisia
e-mail: ridha.bouallegue@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_31,

a good scheme to combat frequency-selective fading channels in wireless communication systems without forgetting the major property that is subcarrier orthogonality.

Under multi path spread situation, a dynamic estimation of channel is necessary before the demodulation of OFDM signals to ensure a coherent detection and since the radio channel is frequency selective and time-varying for wideband mobile communication systems [1, 2].

Channel estimation schemes were categorized according to whether the channel was slowly or rapidly time varying. Generally speaking, we can divide the channel estimation approaches in two major categories, the first in non blind and the second is dedicated for blind estimation. In the earlier one, we usually use training sequences (pilots or preambles) for achieving channel estimation added with standards algorithms such as LS, MMSE LMMSE…

Blind estimation had incited in these last years more confirmed interests by the continuous improvements of estimation results. Hence, for blindly estimate the channel, we can use a subspace approach or we can have recourse to the standards channel estimation algorithms such as Basis Expansion Model or LS, MMSE, LMS or Maximum Likelihood estimators.

In [3], a zero padding SISO-OFDM system associated with either recursive Least Squares or least mean squares method for blind adaptive channel estimation was considered.

In [4], Heath el al., proposed a subspace method using cyclic correlation of the channel output to blindly estimate the channel in OFDM systems. But the estimated channel error in that study is large a subspace approach based on second order statistics is proposed to blindly identify the channel in OFDM systems.

In this paper we will focus on a subspace approach for blindly identify the channel and show the performance of this model comparably with non blind estimators. Section 2 gives an overview of the system model. In Sect. 3, the blind channel estimator based on the subspace approach will be presented, the Sect. 4 provides the experiment results and the conclusion we be held in Sect. 5.

2 System Model

In an OFDM system model, the serial data is converted into M parallel flows. Each parallel data stream modulates a different carrier. The frequency spacing between the adjacent carriers is 1/T, where T is the symbol duration for the parallel data that is M times of the symbol duration for the serial data. Let us consider an OFDM signal in the interval (nT,(n+1)T) as [5]

$$s(t) = \sum_{m=0}^{M-1} a_m(n) e^{jw_m t} \tag{1}$$

where $a_m(n)$are symbols resulting from a modulation constellation as 16 QAM,w_m is the frequency of mth carrier that is $m\frac{2\pi}{T}$. The M samples that are sampled at

$t = nT + i\frac{1}{T},\ i = 0, 1, \ldots, M - 1$ are as follow:

$$s(nM + i) = \sum\nolimits_{m=0}^{M-1} a_m(n)e^{j\frac{2\pi}{M}mi} \tag{2}$$

The M samples can be seen as the inverse discrete Fourier transform (IDFT) of a block for M input symbols.

From the theoretical context, when the number of carriers is great enough, symbol duration T is much larger than the duration of finite impulse response channel; then the inter symbol interference ISI is negligible. However, for the high-bit-rate communications, it is impractical to choose M vey big to make ISI negligible. Thus, a cyclic prefix of length P is added into each block of IDFT output at the transmitter. The length of the prefix is chosen to be longer than the length of the channel impulse response in order to avoid inter-block interference (IBI) [6]. That results with total cancellation of ISI and inter carrier interference (ICI). The input data will be as follow:

$$s(n(M + P) + i) = \sum\nolimits_{m=0}^{M-1} a_m(n)e^{j\frac{2\pi}{M}}(i - p) \tag{3}$$

$$i = 0, 1, \ldots, M + P - 1$$

where $s(n(M + P) + i),\ i = 0, 1, \ldots, P - 1$ denotes the cyclic prefix.

Assuming that the channel is frequency selective affected by an additive white Gaussian noise (AWGN), The received signal r(n) will be degraded by theses two conditions. We will suppose also that the length L of the channel impulse response is known. Assuming that blocks are synchronized and carrier frequency offset is corrected [7], the receiver removes the first P symbols corresponding to the cyclic prefix and performs an M-point DFT on the remaining samples of received signal to obtain $y_i(n), i = 0, 1, \ldots, M - 1$.. If the cyclic prefix duration is equal or more than the channel duration, i.e. $P \geq L$, it is shown that [6, 8]

$$y_i\ (n) = a_i\ (n)\ H\left(\frac{2\pi}{M}i\right) + \nu_i\ (n) \tag{4}$$

where H(.) is the frequency response of the channel.

3 Subspace Approach

We consider a data model for the OFDM system model described in the Sect. 1. Let us indicate the vector $a\ (n) = [a_0\ (n)\ , a_1\ (n)\ , \ldots, a_{M-1}\ (n)]^T$ as the nth block of data and $s\ (n) = [s_{K-1}\ (n)\ , [s_{K-2}\ (n)\ , \ldots, [s_0\ (n)]^T$ as a sequence of the nth block of the IDFT output and embedded cyclic prefix, where K = M + P and P is the length of cyclic prefix.

Denoting $W \triangleq e^{j\frac{2\pi}{M}}, i = 0, 1, \ldots, K-1\ , j = 0, 1, \ldots, M-1$, the nth transmitted data sequence will be written as:

$$s(n) = Wa(n) \tag{5}$$

By considering N blocks of data, $a(n) = [a(n)^T, a(n-1)^T, \ldots, a(n-N+1)^T]^T$, s $= [s(n)^T, s(n-1)^T, \ldots, s(n-N+1)^T]^T$ and $\check{W} = I_N \otimes W$where I_N is an $N \times N$ identity matrix and $\otimes$ is the Kronecker product; the received signal is expressed as:

$$r = Hs + b = H\check{W}a + b \tag{6}$$

With $s = \check{W}a$ and r is an $(NK-L)\times 1$ vector. Let the nth block of received signal be denoted as $r(n) = [r_{K-1}(n), [r_{K-2}(n), \ldots, [r_0(n)]^T$ then $r = [r(n)^T, r(n-1)^T,$ $r(n-N+2)^T,\ r(n-N+1)^T(1:K-L)^T\ ]^T$ where $r(n-N+1)^T(1:K-L)^T$ is a Matlab notation standing for the first K−L elements of $r(n-N+1)$. b is a noise vector that is assumed to be zero mean white Gaussian noise with variance matrix $\sigma^2 I_{NK-L}$ and be mutually independent with the input symbol sequence. H is an (NK−L) × NK matrix defined as [5]:

$$H = \begin{bmatrix} \mathrm{h}_0 & \ldots & \mathrm{h}_L & 0 & & \ldots & 0 \\ 0 & \mathrm{h}_0 & \cdots \mathrm{h}_L & & 0\ \cdots & \cdots & 0 \\ \vdots & \ldots & \ddots & \ldots & & \ddots & \vdots \\ 0 & \ldots & \cdots\ 0 & \mathrm{h}_0 & \ldots & & \mathrm{h}_L \end{bmatrix} \tag{7}$$

Denote $A \triangleq H\check{W}$n then (6) becomes

$$r = Aa + b \tag{8}$$

Since the matrix A should be full column rank for the channel to be identified, we give a sufficient condition for full rank requirement; this will be verified under the assumption $L \leq PN$ [5].

In fact it is the inserted prefix that makes matrix A a "tall" matrix and be possible to be full column rank. The full column rank condition can always be satisfied as long as N is chosen to be large enough.

The identification problem is based on the $(NK-L) \times (NK-L)$ autocorrelation matrix R_r of the measurement where $R_r = E\{rr^H\}$. Considering the expression given in (8), this autocorrelation matrix will be:

$$R_r = AR_aA^H + \sigma^2 I \tag{9}$$

If matrix A is full column rank and the autocorrelation of input R_a, is also full rank, then $range(A)$ =range(AR_aA^H). Let us define the noise subspace for R_r, to be the subspace generated by PN−L eigenvectors correspond-ing to the smallest eigen value, and let $G = [G_1, \ldots, G_{PN-L}]$ be the ma-trix containing those eigenvectors. Then, G spans the null space of $AR_a A^H$ and is orthogonal to its range space:

$$G_i^H A = 0 \quad i = 1, \ldots, PN - L \tag{10}$$

For determining the identifiability of the channel, the equation above will be solved in the least square sense. Let us define ξ_k as:

$$\xi_k = \begin{bmatrix} \mathrm{G}_{k,0} & \cdots & \mathrm{G}_{k,J} & 0 & \ldots & \ldots & 0 \\ 0 & \mathrm{G}_{k,0} & \cdots & \mathrm{G}_{k,J} & 0 \cdots & \cdots & 0 \\ \vdots & \cdots & \ddots & \cdots & \ddots & & \ldots \\ 0 & \ldots & \cdots\ 0 & \mathrm{G}_{k,0} & \cdots & & \mathrm{G}_{k,J} \end{bmatrix} \tag{11}$$

where $J = \mathrm{KN{-}L{-}1}$. Next, $G_k^T H = h^T \xi_k$. This leads to the minimization problem:

$$\begin{aligned} \widehat{h} &= \arg\min \sum\nolimits_{k=1}^{PN-L} \widehat{G}_k^H H \check{W} \check{W}^H H^H \widehat{G}_k \qquad (12) \\ &= \arg\min \sum_{k=1}^{PN-L} h^H \widehat{\xi}_k \check{W} \check{W}^H \xi_k^H h \\ &= \arg\min \sum_{k=1}^{PN-L} h^H \Psi h \end{aligned}$$

where $\Psi = \sum_{k=1}^{PN-L} \widehat{\xi}_k \check{W} \check{W}^H \widehat{\xi}_k^H$. Minimization is subject to the constraint $\|h\| = 1$ Therefore, $\widehat{h}$ is given by the eigenvector corresponding to the smallest eigen value of ψ [5].

4 Experiment Results

Considering the channel model detailed in the second section and for comparing our improved approach with others related works, we will show in this section that the subspace approach for the blind estimation is a robust estimator for a fast time varying channel.

The simulations was carried for a 16 QAM modulation even if it can be achieved by the others schemes as BPSK, QPSK,…

The performance of the estimator is evaluated in term of bit error rate BER versus the signal to noise ratio SNR and in term of mean square error versus SNR.

The Fig. 1 shows the behavior of the channel:

Form this figure, we mainly remark that the mean square error for our simulation is low comparably with those given in related works as given in [9]. In fact, comparing

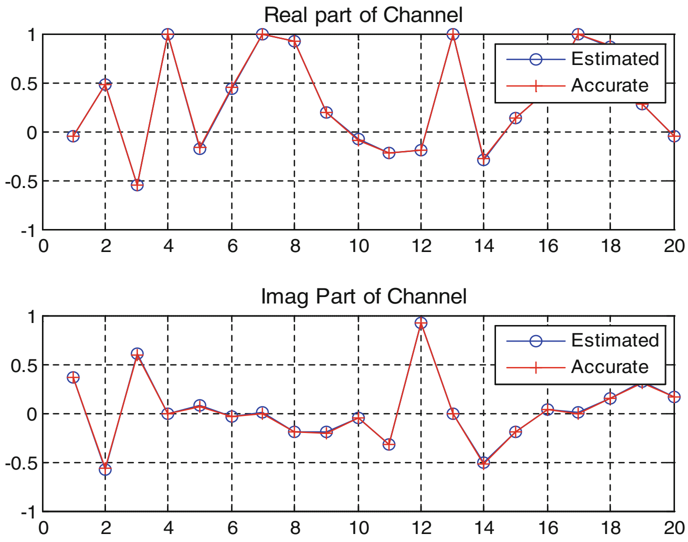


Fig. 1 Channel behavior under a 16 QAM modulation

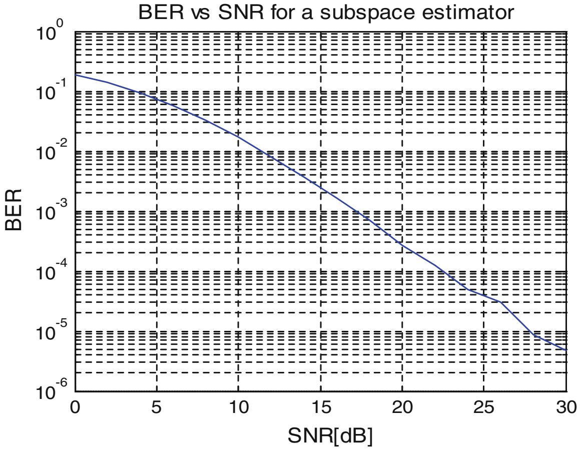


Fig. 2 BER versus SNR for a subspace estimator

the performances of our simulations with many related works, we denote at least a gain of more than 1.5 dB which proves the effectiveness of our work (Fig. 2).

The Fig. 3 behind shows the bias of the channel; it is worthwhile that is an acceptable behavior and transmitted and received signals will not be very affected by the variation of the channel.

Figure 4 shows the plot of the MSE versus the SNR; starting from the results, it is clear that our approach performs well because the mean square error is low.

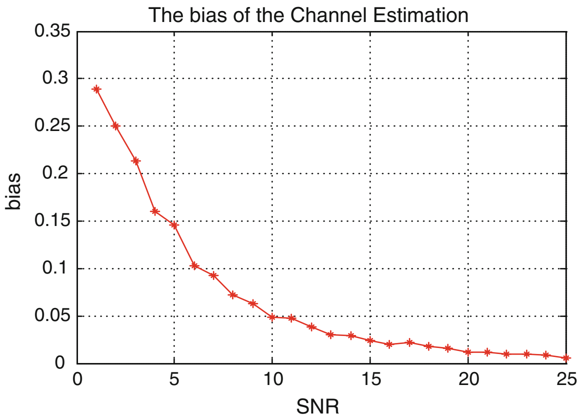


Fig. 3 The biais of the channel versus SNR

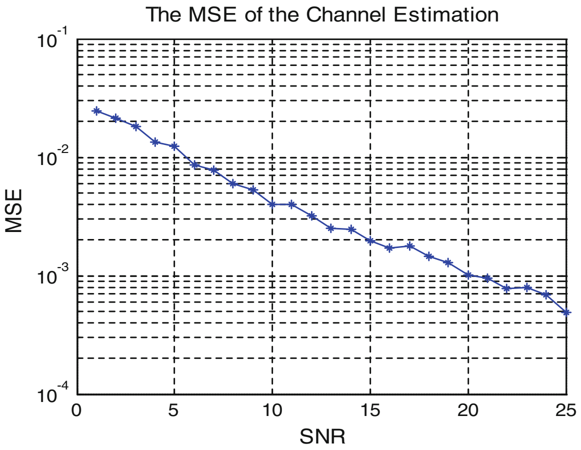


Fig. 4 MSE versus SNR for a subspace estimator

5 Conclusion

In this article, a subspace approach was improved for blindly identify and estimate a fast time varying channel. It is confirmed from related works that this approach is more complex computational estimator than the others methods based in the training sequences.

But from the results of our improved simulations, we can prove that we the channel wins more in term of spectral efficiency and in term of reduction of the bit error rate and the mean square error.

This approach will be extended in a MIMO context to brush up our work.

References

1. Chotikakamthorn N, Suzuki H (1999) On Identifiability of OFDM Blind Channel Estimation, 0–7803-5435-4/99/ 1999 IEEE
2. Bahai ARS, Saltzberg BR (1999) Multi-carrier digital communications: theory and applications of OFDM. Kluwer Academic/Plenum
3. Doukopoulos XG, Moustakides GV (2006) Blind adaptive channel estimation in OFDM systems. IEEE Trans Wirel Commun 5(7):1716–1725
4. Heath RW, Giannakis GB (1999) Exploiting input cyclostationary for blind channel identification in OFDM systems. IEEE l"s. Signal Process 47(3):848–856
5. Cai X, Akansu AN (2000) A subspace method for blind channel identification in OFDM systems, 0–7803-6283-7/00/ 0 2000 IEEE
6. Zaier A, Bouallegue R (2011) A full performance analysis of channel estimation methods for time VARYING OFDM systems, STA Conference 2011
7. van de Beek J-J, Sandell M, Borjesson P0, (1997) ML estimation of time and frequency offset in OFDM systems. IEEE Trans Signal Process 45:1800–1805
8. Cioffi JM (1991) "A multicarrier primer", Amatio Communications Corporation and Stanford University
9. Necker MC (2004) Stuber GL Totally Blind Channel Estimation for OFDM on Fast Varying Mobile Radio Channels

Adaptive Control and Synchronization Design for the Lu-Xiao Chaotic System

Vaidyanathan Sundarapandian

Abstract This paper derives new results on the adaptive control and synchronization design for the Lu-Xiao chaotic system (2012) with unknown parameters. First, adaptive control is designed to stabilize the Lu-Xiao chaotic system to its unstable equilibrium at the origin. Then adaptive synchronization is designed to achieve global chaos synchronization of identical Lu-Xiao chaotic systems with unknown parameters. The adaptive control and synchronization results derived in this paper have been established using Lyapunov stability theory. Numerical simulations are given to validate and depict the effectiveness of the adaptive control and synchronization laws derived in this paper for the Lu-Xiao chaotic system.

1 Introduction

Chaotic systems are nonlinear dynamical systems which have some special features such as being extremely sensitive to small variations of initial conditions. The chaos phenomenon was first observed in weather models by Lorenz ([1], 1963). This was followed by a discovery of a large number of chaotic systems in Physics, Chemistry, Biology and Engineering [2].

The problem of controlling a chaotic system was introduced by Ott et al. ([3], 1990). The control of chaotic systems is basically to design state feedback control laws that stabilizes the chaotic systems around the unstable equilibrium points.

Active control method is used when the system parameters are known. Adaptive control method is used when some or all of the system parameters are unknown [3–5] and the controller makes use of estimates of the unknown parameters for implementation.

V. Sundarapandian (✉)
Research and Development Centre, Vel Tech Dr. RR & Dr. SR Technical University,
Avadi, Chennai, Tamil Nadu 600 062, India
e-mail: sundarvtu@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_32,

Since the pioneering work by Pecora and Carroll ([6], 1990), several approaches have been proposed for chaos synchronization such as the active control method [7, 8], adaptive control method [9], sampled-data control method [10], backstepping method [11], sliding mode control method [12, 13], etc.

This paper derives new results for the adaptive control and synchronization design for the Lu-Xiao chaotic system (Lu and Xiao, [14], 2012). First, we devise adaptive control scheme using state feedback control for the Lu-Xiao system about its unstable equilibrium at the origin. Then we devise adaptive synchronization scheme for identical Lu-Xiao systems. The adaptive control and synchronization results derived in this paper are established using Lyapunov stability theory.

2 Adaptive Control Design of the Lu-Xiao Chaotic System

In this section, we discuss the design of adaptive controller for globally stabilizing the Lu-Xiao system (2012), when the parameter values are unknown.

Thus, we consider the controlled Lu-Xiao system described by the dynamics

$$\begin{aligned} \dot{x}_1 &= a(x_2 - x_1) + x_2 x_3 + u_1 \\ \dot{x}_2 &= -b x_1 x_3 + c x_1 + u_2 \\ \dot{x}_3 &= d x_1 x_2 - \varepsilon x_3 + u_3 \end{aligned} \tag{1}$$

where x_1, x_2, x_3 are the states and u_1, u_2, u_3 are the controls.

Next, we consider the following feedback control functions

$$\begin{aligned} u_1 &= -\hat{a}(x_2 - x_1) - x_2 x_3 - k_1 x_1 \\ u_2 &= \hat{b} x_1 x_3 - \hat{c} x_1 - k_2 x_2 \\ u_3 &= -\hat{d} x_1 x_2 + \hat{\varepsilon} x_3 - k_3 x_3 \end{aligned} \tag{2}$$

where $\hat{a}, \hat{b}, \hat{c}, \hat{d}, \hat{\varepsilon}$ are estimates of the system parameters a, b, c, d, ε, respectively, and k_i, $(i = 1, 2, 3)$ are positive constants.

Substituting the control law (2) into the plant equation (1), we obtain

$$\begin{aligned} \dot{x}_1 &= (a - \hat{a})(x_2 - x_1) - k_1 x_1 \\ \dot{x}_2 &= -(b - \hat{b}) x_1 x_3 + (c - \hat{c}) x_1 - k_2 x_2 \\ \dot{x}_3 &= (d - \hat{d}) x_1 x_2 - (\varepsilon - \hat{\varepsilon}) x_3 - k_3 x_3 \end{aligned} \tag{3}$$

We define the parameter estimation error as

$$e_a = a - \hat{a}, \quad e_b = b - \hat{b}, \quad e_c = c - \hat{c}, \quad e_d = d - \hat{d}, \quad e_\varepsilon = \varepsilon - \hat{\varepsilon} \tag{4}$$

Using (4), the state dynamics (3) can be simplified as

$$\begin{aligned} \dot{x}_1 &= e_a(x_2 - x_1) - k_1 x_1 \\ \dot{x}_2 &= -e_b x_1 x_3 + e_c x_1 - k_2 x_2 \\ \dot{x}_3 &= e_d x_1 x_2 - e_\varepsilon x_3 - k_3 x_3 \end{aligned} \tag{5}$$

Consider the quadratic Lyapunov function defined by

$$V = \frac{1}{2}\left(x_1^2 + x_2^2 + x_3^2 + e_a^2 + e_b^2 + e_c^2 + e_d^2 + e_\varepsilon^2\right) \tag{6}$$

which is a positive definite function on $\mathbb{R}^8$.

Note that

$$\dot{e}_a = -\dot{\hat{a}}, \quad \dot{e}_b = -\dot{\hat{b}}, \quad \dot{e}_c = -\dot{\hat{c}}, \quad \dot{e}_d = -\dot{\hat{d}}, \quad \dot{e}_\varepsilon = -\dot{\hat{\varepsilon}} \tag{7}$$

Differentiating V along the trajectories of (5) and using (7), we obtain

$$\begin{aligned} \dot{V} = -k_1 x_1^2 - k_2 x_2^2 - k_3 x_3^2 + e_a\left[x_1(x_2 - x_1) - \dot{\hat{a}}\right] + e_b\left[-x_1 x_2 x_3 - \dot{\hat{b}}\right] \\ + e_c\left[x_1 x_2 - \dot{\hat{c}}\right] + e_d\left[x_1 x_2 x_3 - \dot{\hat{d}}\right] + e_\varepsilon\left[-x_3^2 - \dot{\hat{\varepsilon}}\right] \end{aligned} \tag{8}$$

In view of Eq. (8), the estimated parameters are updated by the following law:

$$\begin{aligned} \dot{\hat{a}} &= x_1(x_2 - x_1) + k_4 e_a \\ \dot{\hat{b}} &= -x_1 x_2 x_3 + k_5 e_b \\ \dot{\hat{c}} &= x_1 x_2 + k_6 e_c \\ \dot{\hat{d}} &= x_1 x_2 x_3 + k_7 e_d \\ \dot{\hat{\varepsilon}} &= -x_3^2 + k_8 e_\varepsilon \end{aligned} \tag{9}$$

where k_i, $(i = 4, \ldots, 8)$ are positive constants.

Theorem 1 *The Lu-Xiao chaotic system* (1) *with unknown parameters is globally and exponentially stabilized by the adaptive control law* (2)*, where the update law for the parameters is given by* (9) *and* k_i, $(i = 1, 2, \ldots, 8)$ *are positive constants.*

Proof Substituting (9) into (8), we obtain

$$\dot{V} = -k_1 x_1^2 - k_2 x_2^2 - k_3 x_3^2 - k_4 e_a^2 - k_5 e_b^2 - k_6 e_c^2 - k_7 e_d^2 - k_8 e_\varepsilon^2 \tag{10}$$

which is a negative definite function on $\mathbb{R}^8$.

Thus, by Lyapunov stability theory [15], the plant dynamics (5) is globally exponentially stable and also that the parameter estimation errors $e_a, e_b, e_c, e_d, e_\varepsilon$ converge to zero exponentially with time. □

For simulations, the fourth order Runge-Kutta method with step-size $h = 10^{-8}$ is used to solve the Lu-Xiao system (1) with the adaptive control law (2) and the parameter update law (9). We take the gains as $k_i = 4$ for $i = 1, 2, \ldots, 8$.

The parameters of the system (1) are selected as

$$a = 20, \quad b = 5, \quad c = 40, \quad d = 4 \quad \text{and} \quad \varepsilon = 3$$

Suppose that the initial values of the estimated parameters are

$$\hat{a}(0) = 4, \quad \hat{b}(0) = 12, \quad \hat{c}(0) = 10, \quad \hat{d}(0) = 20, \quad \hat{\varepsilon}(0) = 8$$

Suppose that we take the initial values of the states of (1) as

$$x_1(0) = 16, \quad x_2(0) = -25, \quad x_3(0) = 14$$

We find that the controlled Lu-Xiao system (1) converges to $E_0 = (0, 0, 0)$ exponentially as shown in Fig. 1 and the parameter estimation errors $e_a, e_b, e_c, e_d, e_\varepsilon$ converge exponentially to zero as shown in Fig. 2.

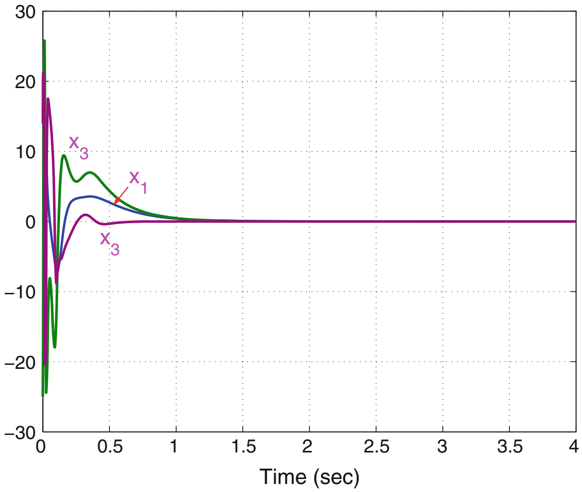


Fig. 1 Time responses of the controlled Lu-Xiao system

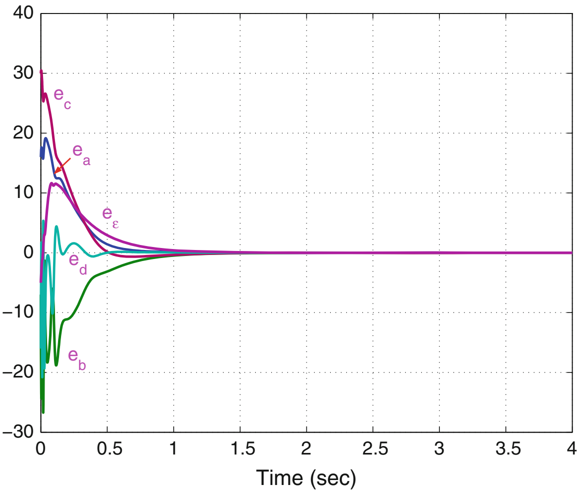


Fig. 2 Time-history of the parameter estimation errors $e_a, e_b, e_c, e_d, e_\varepsilon$

3 Adaptive Synchronization Design for the Lu-Xiao Chaotic Systems

In this section, we describe the design of adaptive synchronizer for identical Lu-Xiao systems (2012) with unknown parameters.

As the master system, we consider the Lu-Xiao dynamics described by

$$\begin{aligned} \dot{x}_1 &= a(x_2 - x_1) + x_2 x_3 \\ \dot{x}_2 &= -b x_1 x_3 + c x_1 \\ \dot{x}_3 &= d x_1 x_2 - \varepsilon x_3 \end{aligned} \tag{11}$$

where x_1, x_2, x_3 are the state variables and a, b, c, d, ε are unknown system parameters.

As the slave system, we consider the controlled Lu-Xiao dynamics described by

$$\begin{aligned} \dot{y}_1 &= a(y_2 - y_1) + y_2 y_3 + u_1 \\ \dot{y}_2 &= -b y_1 y_3 + c y_1 + u_2 \\ \dot{y}_3 &= d y_1 y_2 - \varepsilon y_3 + u_3 \end{aligned} \tag{12}$$

where y_1, y_2, y_3 are the state variables and u_1, u_2, u_3 are the nonlinear controllers to be designed.

The synchronization error e is defined by

$$e_i = y_i - x_i, \quad (i = 1, 2, 3) \tag{13}$$

Then the error dynamics is obtained as

$$\begin{aligned} \dot{e}_1 &= a(e_2 - e_1) + y_2 y_3 - x_2 x_3 + u_1 \\ \dot{e}_2 &= c e_1 - b(y_1 y_3 - x_1 x_3) + u_2 \\ \dot{e}_3 &= -\varepsilon e_3 + d(y_1 y_2 - x_1 x_2) + u_3 \end{aligned} \tag{14}$$

We define the adaptive synchronizing law

$$\begin{aligned} u_1 &= -\hat{a}(e_2 - e_1) - y_2 y_3 + x_2 x_3 - k_1 e_1 \\ u_2 &= -\hat{c} e_1 + \hat{b}(y_1 y_3 - x_1 x_3) - k_2 e_2 \\ u_3 &= \hat{\varepsilon} e_3 - \hat{d}(y_1 y_2 - x_1 x_2) - k_3 e_3 \end{aligned} \tag{15}$$

where $\hat{a}, \hat{b}, \hat{c}, \hat{d}, \hat{\varepsilon}$ are estimates of the system parameters a, b, c, d, ε, respectively, and k_i, $(i = 1, 2, 3)$ are positive constants.

Substituting (15) into (14), we obtain the error dynamics as

$$\begin{aligned} \dot{e}_1 &= (a - \hat{a})(e_2 - e_1) - k_1 e_1 \\ \dot{e}_2 &= (c - \hat{c}) e_1 - (b - \hat{b})(y_1 y_3 - x_1 x_3) - k_2 e_2 \\ \dot{e}_3 &= -(\varepsilon - \hat{\varepsilon}) e_3 + (d - \hat{d})(y_1 y_2 - x_1 x_2) - k_3 e_3 \end{aligned} \tag{16}$$

We define the parameter estimation error as

$$e_a = a - \hat{a}, \quad e_b = b - \hat{b}, \quad e_c = c - \hat{c}, \quad e_d = d - \hat{d}, \quad e_\varepsilon = \varepsilon - \hat{\varepsilon} \tag{17}$$

Substituting (17) into (16), the error dynamics (16) can be simplified as

$$\begin{aligned} \dot{e}_1 &= e_a(e_2 - e_1) - k_1 e_1 \\ \dot{e}_2 &= e_c e_1 - e_b(y_1 y_3 - x_1 x_3) - k_2 e_2 \\ \dot{e}_3 &= -e_\varepsilon e_3 + e_d(y_1 y_2 - x_1 x_2) - k_3 e_3 \end{aligned} \tag{18}$$

Consider the quadratic Lyapunov function defined by

$$V = \frac{1}{2}(e_1^2 + e_2^2 + e_3^2 + e_a^2 + e_b^2 + e_c^2 + e_d^2 + e_\varepsilon^2) \tag{19}$$

which is a positive definite function on $\mathbb{R}^8$.

Note that

$$\dot{e}_a = -\dot{\hat{a}}, \quad \dot{e}_b = -\dot{\hat{b}}, \quad \dot{e}_c = -\dot{\hat{c}}, \quad \dot{e}_d = -\dot{\hat{d}}, \quad \dot{e}_\varepsilon = -\dot{\hat{\varepsilon}} \tag{20}$$

Differentiating V along the trajectories of (18) and using (20), we obtain

$$\begin{aligned} \dot{V} = &- k_1 e_1^2 - k_2 e_2^2 - k_3 e_3^2 + e_a \left[e_1(e_2 - e_1) - \dot{\hat{a}} \right] + e_b \left[-e_2(y_1 y_3 - x_1 x_3) - \dot{\hat{b}} \right] \\ &+ e_c \left[e_1 e_2 - \dot{\hat{c}} \right] + e_d \left[e_3(y_1 y_2 - x_1 x_2) - \dot{\hat{b}} \right] + e_\varepsilon \left[-e_3^2 - \dot{\hat{\varepsilon}} \right] \end{aligned} \tag{21}$$

In view of Eq. (21), the estimated parameters are updated by the following law:

$$\begin{aligned}
\dot{\hat{a}} &= e_1(e_2 - e_1) + k_4 e_a \\
\dot{\hat{b}} &= -e_2(y_1 y_3 - x_1 x_3) + k_5 e_b \\
\dot{\hat{c}} &= e_1 e_2 + k_6 e_c \\
\dot{\hat{d}} &= e_3(y_1 y_2 - x_1 x_2) + k_7 e_d \\
\dot{\hat{\varepsilon}} &= -e_3^2 + k_8 e_\varepsilon
\end{aligned} \tag{22}$$

where k_i, $(i = 4, \ldots, 8)$ are positive constants.

Theorem 2 *The identical Lu-Xiao chaotic systems* (11) *and* (12) *with unknown parameters are globally and exponentially synchronized by the adaptive control law* (15)*, where the update law for the parameters is given by* (22) *and* k_i, $(i = 1, 2, \ldots, 8)$ *are positive constants.*

Proof Substituting (22) into (21), we obtain

$$\dot{V} = -k_1 x_1^2 - k_2 x_2^2 - k_3 x_3^2 - k_4 e_a^2 - k_5 e_b^2 - k_6 e_c^2 \tag{23}$$

which is a negative definite function on $\mathbb{R}^8$.

Thus, by Lyapunov stability theory [15], it follows that the error dynamics (18) is globally exponentially stable and also that the parameter estimation errors $e_a, e_b, e_c, e_d, e_\varepsilon$ converge to zero exponentially as $t \to \infty$. This completes the proof. □

For the numerical simulations, the fourth order Runge-Kutta method with step-size $h = 10^{-6}$ is used to solve the Lu-Xiao systems (11) and (12) with the adaptive control law (15) and the parameter update law (22). We take the gains as $k_i = 4$ for $i = 1, 2, \ldots, 8$.

The parameters of the Lu-Xiao system (11) are selected as

$$a = 20, \quad b = 5, \quad c = 40, \quad d = 4 \quad \text{and} \quad \varepsilon = 3$$

Suppose that the initial values of the estimated parameters are

$$\hat{a}(0) = 12, \quad \hat{b}(0) = 36, \quad \hat{c}(0) = 1, \quad \hat{d}(0) = 41, \quad \hat{\varepsilon}(0) = 5$$

Suppose that the initial values of the master system (11) are taken as

$$x_1(0) = 17, \quad x_2(0) = -15, \quad x_3(0) = 28$$

Suppose that the initial values of the slave system (12) are taken as

$$y_1(0) = 24, \quad y_2(0) = 18, \quad y_3(0) = -9$$

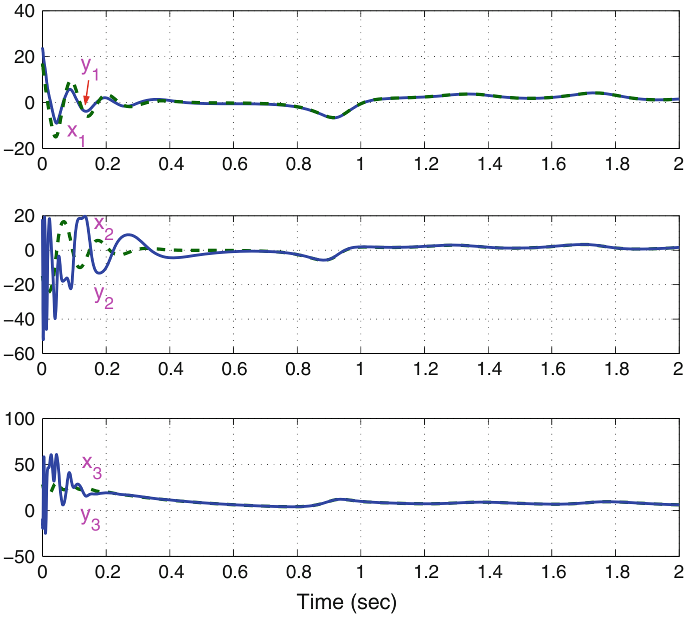


Fig. 3 Synchronization of the Lu-Xiao chaotic systems

Figure 3 depicts the synchronization of identical Lu-Xiao systems.

Figure 4 depicts the exponential convergence of the parameter estimation errors $e_a, e_b, e_c, e_d, e_\varepsilon$ to zero.

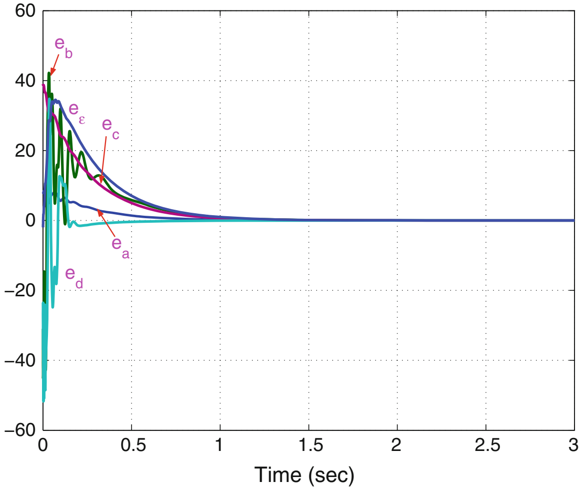


Fig. 4 Time-history of the parameter estimation errors $e_a, e_b, e_c, e_d, e_\varepsilon$

4 Conclusions

In this paper, we derived results for the adaptive controller and synchronizer for the Lu-Xiao chaotic system [14] with unknown parameters. First, we designed adaptive control law to stabilize the Lu-Xiao system to its unstable equilibrium point at the origin based on the Lyapunov stability theory. Then we designed adaptive synchronizer for the global chaos synchronization of identical Lu-Xiao systems with unknown parameters. Our synchronization results were established using Lyapunov stability theory. Numerical simulations are presented to illustrate the effectiveness of the proposed adaptive controller and synchronizer schemes for the Lu-Xiao chaotic system (2012).

References

1. Lorenz EN (1963) Deterministic nonperiodic flow. J Atmos Sci 20(2):130–141
2. Lakshmanan M, Murali K (1996) Nonlinear oscillators: controlling and synchronization. World Scientific, Singapore
3. Ott E, Grebogi C, Yorke JA (1990) Controlling chaos. Phys Rev Lett 64:1196–1199
4. Ge SS, Wang C, Lee TH (2000) Adaptive backstepping control of a class of chaotic systems. Int J Bifurcat Chaos 10:1149–1156
5. Sun M, Tian L, Jiang S, Xun J (2007) Feedback control and adaptive control of the energy resource chaotic system. Chaos, Solitons & Fractals 32:168–180
6. Pecora LM, Carroll TL (1990) Synchronization in chaotic systems. Phys Rev Lett 64:821–824
7. Lu L, Zhang C, Guo ZA (2007) Synchronization between two different chaotic systems with nonlinear feedback control. Chin Phys 16(6):1603–1607
8. Sundarapandian V (2011) Hybrid chaos synchronization of hyperchaotic Liu and hyperchaotic Chen systems by active nonlinear control. Int J Comput Sci Eng Inf Technol 1(2):1–14
9. Liao TL, Tsai SH (2000) Adaptive synchronization of chaotic systems and its applications to secure communications. Chaos, Solitons & Fractals 11:1387–1396
10. Yang T, Chua LO (1999) Control of chaos using sampled-data feedback control. Int J Bifurcat Chaos 9:215–219
11. Yu YG, Zhang SC (2006) Adaptive backstepping synchronization of uncertain chaotic systems. Chaos, Solitons & Fractals 27:1369–1375
12. Konishi K, Hirai M, Kokame H (1998) Sliding mode control for a class of chaotic systems. Phys Lett A 245:511–517
13. Sundarapandian V (2011) Global chaos synchronization of Pehlivan systems by sliding mode control. Int J Comput Sci Eng 3(5):2163–2169
14. Lu H, Xiao X (2012) Analysis of a novel autonomous chaotic system. Int J Adv Comput Technol 4(1):248–255
15. Hahn W (1967) The stability of motion. Springer, New York

Part V
The Fourth International Conference on Networks & Communications (NETCOM-2012): Network Operations and Management

Secure Patient Monitoring and Self-management Using Brain Expression Interpreter

Suleyman Kondakci and Dilek Doruk

Abstract This article presents a brief description of an experimental system designed for self-managing and monitoring of patients having difficulties in verbal communication. A vital role of the proposed system is to engage individuals to be effective self-managers of their own health. Technology-based interventions to support self-management across the spectrum of chronic diseases have the potential to reach a broader population of patients. We believe that studies featuring extensive research of self-management of patients with different chronic conditions are needed to explore the interaction of technology-based interventions.

1 Introduction

We present here a framework consisting of design and implementation of a novel system that effectively enhances quality of medical care for patients with reduced communication abilities. Many physically reduced patients have difficulties in expressing their basic requirements, claims, sufferings, personal needs, and achieving the most basic environmental/clinical comforts. These patients can still use cognitive and expressive skills to convey their needs to physicians and manage part of their own medical treatments. Results of several simulations and experimentations with the implemented system have shown significant improvements in responses to patient requests and quality of the related patient care.

S. Kondakci (✉)
Faculty of Engineering and Computer Sciences,
Izmir University of Economics, 35330 Balcova-Izmir, Turkey
e-mail: suleyman.kondakci@ieu.edu.tr

D. Doruk
Asım Arar Vocational Health School, Tire-zmir, Turkey
e-mail: dilekdoruk72@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_33,

Many of us can successfully use the nonverbal communication in the form of gestures and interpretation of EEG/EMG signals in order to express most of our feelings. Combining this type of communication with the emerging technologies, we can substantially increase the quality of care for many bedridden patients. In addition to guiding diagnosis, management, and treatment of patients for healthcare providers, the system can be used to enhance the comfort for both patients and medical staff, raise quality of care, reduce several kinds of risk to patients and to healthcare providers.

The proposed system, called Brain Expression Interpreter (BEI), improves the quality of care and self-management of patients by a computerized system installed on a patient bed and on the health provider's server. The system communicates with the patient using a headset of electrodes worn by the patient. Although the system requires patient training prior to using it, it can be very useful for long term hospitalizations. Also the cost of the total care can be drastically reduced by the decentralizing the care staff, while managing the most basic needs of the patient locally. Many practical healthcare functions/services can be automatically managed by use of BEI. These services include management of the natural needs of patients, contactless control of the patient bed, messaging with relatives, Internet, and remote control of home appliances. Sensors attached to the patient bed can collect vital body information and environment data in order to automatically monitor and control patient discomforts. A functionality of BEI is that the patient can express his/her situation by use of a speech synthesizer. The patient can create any text by use of pictograms, e.g., "call my daughter", which can then be converted to a real-time speech.

In addition to active patient monitoring, engaging patients in the self-management (especially bedridden patients) is one of the major application area of this configurable intelligent system. Independent of the location of medical staff, a patient can be remotely monitored and guided to self-care. For example, a sudden attack (e.g., an epileptic seizure) can be immediately reported to remotely located medical staff. Several of the basic services provided manually by the care staff can also be performed via the interaction of patients' thoughts through BEI. For example, a bedridden person can command BEI for adjusting the bed, turning the light on/off, watching TV, listening to music, controlling an air conditioner,peripheral support equipments, and other home appliances.

Chronic patients over time often acquire a high level of knowledge about the processes involved in their own care, and often develop a routine in coping with their conditions. For this class of patients, the proposed system can be a valuable tool to benefit from. Patients can configure the system to adopt their own routine treatment for increased comfortability. Use of such a system for remote patient care can also reduce the cost of expensive visits of patients to hospitals and hiring local care staff. Both patients and healthcare centers are location independent, so that patients can be ubiquitously monitored, while self-managing most of the daily needs. Use of technology in the area of remote health monitoring is rapidly developing, however, impacts of non-standardized application and communication techniques are being hardly discussed. This issue is partly considered in [1]. Recently, most pervasive applications consider the biomedical data exchange, [2], rather than the provision of real-time self-management of patients. One of the main objectives of BEI is aimed at the self-management of patients, while actively engaging care staff remotely.

2 System Overview

In the following, we present a brief overview of how BEI works, and explain how physicians can configure the system for different types of patient treatments. The proposed system supports the use of Internet and mobile devices in providing a role-based healthcare using nonverbal patient communication. The system provides healthcare information to practitioners, real-time monitoring of patient vitals, control of environmental devices, and self-management of daily needs. Three different roles are defined, ubiquitously located care staff, patients, and environmental devices (home appliances) controlled by patients' cognitive/expressive activities. A different role-based mobile care system applied to monitoring of hypertension and arrhythmia patients is also proposed by [3]. A typical network of BEI is shown in Fig. 1, which is composed of a decentralized client/server architecture, where the server system runs in a hospital network to collect and securely distribute patient requests to care units. Patient requests are collected by the BEI client (patient agent) communicating directly with a headset (EMG electrodes) worn by the patient. That is, the patient agents (PAs) are used as a bridge between the patient and care staff, and between the patient and the devices that provide self-management tasks to the patient, e.g., adjusting the patient bed and controlling home appliances.

Functions of PAs are defined in two groups, predefined (default) and adaptive, respectively. Predefined functions guide both the patients and local care staff for

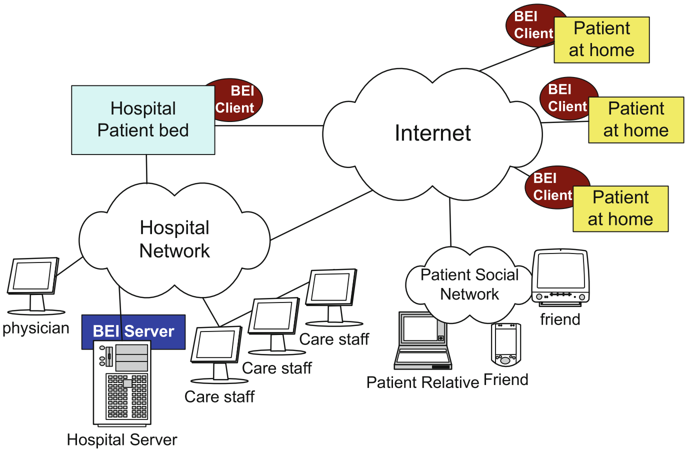


Fig. 1 The communication environment of BEI

the management of patients' natural and environmental needs. Adaptive functions are defined by the respective care center (or doctor) to match the patient-dependent treatments. It should be noted that both the predefined and adaptive PA functions can be expanded to construct more patient requests for designated care staff and self-management tasks as many as needed.

PAs gather EMG signals (cognitive and expressive) and construct associated instructions for both care staff and devices used for self-management. That is, the main task of a PA is to construct instructions activated from brain signals and send the instructions to related destinations. A PA instruction is an encoding (interpretation) of a specific brain signal, either cognitive or expressive, which is used by the care staff or by a remotely controlled home device. As a simple example, knitting of eye-brows encode into an expressive instruction indicating that the patient needs to be taken to toilet. An eye-rolling, such as rotating the eyes upward and downward with accurately captured expressive signals, can be encoded to indicate various states of a patient such as incredulity, contempt, boredom, pain frustration, natural need, or exasperation.

Targets of patient requests are also grouped into two, remote and on-site. Only the adaptive requests collected from a PA are sent to care centers (or practitioners), whereas the predefined requests are used to automatically control the self-management units, e.g., patient bed and other environmental devices. The predefined functions contains sub-tasks that are used to construct and send on-site requests to the care staff (or relatives) local to the patient.

3 System Configuration

Before the system is activated, it must be properly configured by the physician to match the specific patient needs. Configuration means that the physician defines a set of instructions by the aid of the cognitive and expressive signal sets provided by the BEI signal and instruction libraries. Some instructions are sent to care units in audible and textual forms, while other instructions are formated and sent via wireless units to control the self-management equipments. BEI contains predefined default settings for dealing mostly with basic patient needs both environment-wise and care-wise.

Signals are clearly discretized by a specifically defined encoding scheme, so that each expression is mapped to a distinct code called *instruction*. The code, in turn, is mapped to a short textual description (called *interpretation*), which can be easily perceived both by the care staff and the patient when appeared on the PA screen. A single expressive signal can be encoded into a single instruction, or a combination of multiple expressive signals can be used to encode a single instruction depending on the complexity of treatment and patient requirements. For example, raising eyebrow followed by blinking can be encoded into a single instruction. By this way, we can create as many instructions as needed. BEI is an intelligent and reconfigurable system that can be used to freely construct patient-dependent care requests and self management functions. During the initial setup, the physician builds his/her own instruction set by just dragging and combining pictograms. This process is called *task definition*, which is a mandatory task before the system can be activated. Following the task definition, the physician assigns his/her own textual definition (i.e., interpretation) to each task defined so far.

As known, human gestures inherently activate specific expressive signals of the brain, which in turn, trigger dedicated instructions at PAs. For example, the expression shown on the lower right end of Fig. 2 is encoded as "Champing", which indicates pain. Thus, during the system setup, the physician configures BEI with $Champing \Rightarrow Pain$, i.e. champing of this patient now will indicate pain. The expression shown on the lower left end of the figure can be used to indicate anxiety. Encoding and interpretations of the gestures shown in Fig. 2 have been experimented with a very low misinterpretation rate.

Though the use of modern technology can enhance the quality of our lives in many occasions, engaging patients in the self-treatment can be quite challenging. Prior to using BEI, patients are trained to use the system. Especially, it is aimed at reducing the training threshold by use of pictograms for training patients' expressive skills. Figure 2 shows some gestures used for patient training. Besides, during the patient care, PA screens continuously display specifically designed pictographs of gestures to facilitate the communication of patients with PAs, and thus, with care units. Because pictograms can be used to easily and quickly express requests that can be associated with instructions in order to conduct the necessary treatment.

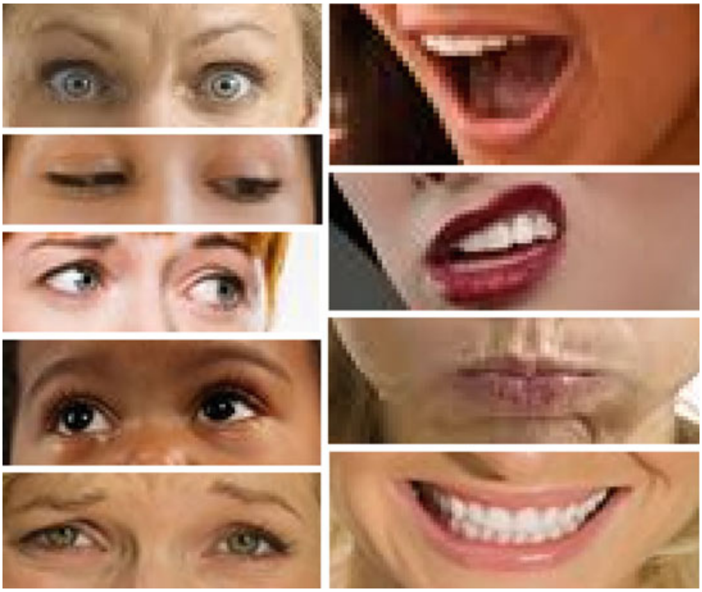


Fig. 2 Typical gestures used for patient training

Systems used within the medical support are subject to pose highest security and user–friendliness, which are also the main constraint for patients aimed at using BEI. The following steps summarize the initiation and use of BEI.

(1) Patient authentication. The main concern has to do with the confidentiality of the patient data transmitted over insecure communication infrastructure such as the Internet. Since the patient is assumed to communicate with remote care staff over insecure networks, a secure exchange of electronic patient record should be provided. BEI provides two modes of communication, secure and insecure (but reliable). Following the initial authentication of patients to practitioners, the remaining communication (e.g., nurse call) can be carried out using insecure mode, which is much more faster. Privacy and data protection issues regarding e-Health have been considered by various work, e.g., [4–6].

(2) Patient training. First of all, ability and willingness of the patient to participate in the collaboration should be verified. Following the patient verification, a common training using pictographs is conducted. In order to increase the accuracy and operational reliability, the patient is first trained by experimenting with predefined pictograms serving as an additional communication means. Accuracy of the error-free interpretation of patient expressions has a central importance. Therefore, the training process may take days, depending on the patient. However, most pictograms used for the environmental control (e.g., adjusting the patient bed) have most frequently been interpreted correctly, i.e., with up to 5 % error rate.

(3) System training. Following a successful patient training, the patient agent will be trained to correctly aggregate and interpret the instructions generated by patient expressions. System training for matching individual patient behavior is a secondary training phase needed to fine-tune the system accuracy so that reliable interpretations of patient claims can be aggregated.

4 Data Exchange Format

To standardize the exchange of information, various coding schemes may be used in combination with internationally recognized medical standards. Health Informatics of the European Committee for Standardisation (CEN/TC 251) prepares standards to exchange electronic health record (EHR) and other measurement data between computer systems. Especially, CEN 13606 EHR communication standard has been undergoing steady revision during the past decade. CEN 13606 EHR contains a five part EHR communication standard. Where, CEN 13606 Part 4 (Security) defines the methodology for specifying the privileges that are necessary to access the EHR data. CEN 13606 Part 5 (Exchange models) describes the messaging model to enable the exchange of EHR data. A version of international standards regarding health informatics is also given in [7]. Since these standards are not at a level to be fully implementable yet, we defined easily portable data formats. BEI uses mainly two types of data exchange processes, *Remote Communication* and *Device Control*. The Remote Communication process deals with the data exchange between the patient node (PA) and remotely (e.g., Internet) located care units, whereas the Device Control processes deal with data exchange between the PA and self-management controller and local devices (e.g., air conditioner, body sensors, environment sensors, patient bed adjustment, etc) that are controlled directly by patient requests employing dedicated gestures.

A common example of Internet data exchange takes place between a patient and a remotely located care staff/unit during the authentication of the patient and also during the private EHR exchange. In these cases, a detailed patient record is sent to the care center. During the Internet communication some part of the data are kept confidential. As the basic requirement, confidential data exchange requires secure communication, whereas controlling of local devices do not require secure communication. Security is provided by encrypting the confidential data using an encryption algorithm and a pair of encryption key. The care center (e.g., physician or hospital) has a copy of the public key of the patient, while the patient has the public key of the remote site and the private key of himself/herself. An example of an Internet exchange (requiring secure mode) is such that a patient sending a photo of a healing wound by email to the physician for control, or sending a psychiatry report of the patient. Another example of an Internet data exchange is when a patient on vacation visits a doctor who then may request access to the patient's health records, such as medicine prescriptions, x-ray photographs, or other test results residing on a remote hospital server. Some security issues regarding electronic health and patient privacy are also considered in [8–10].

With BEI, we focus more on the efficiency and practical use of security transparent to both the patient and the care staff. Figure 3 shows the data packet format involved in the communication of BEI nodes. Two flags are used, S-mod flag for secure mode communication, and Urgent flag for informing urgent situations. The Urgent flag can be used to alert both serious health conditions and system malfunctions. Fields such as patient ID (P-ID), hospital-ID (H-ID), and location of the patient P-location, digital signature (DS) of the patient, and Data field are used for remote communication. In the secure mode, contents of P-ID, H-ID, and P-location fields are used to compute a one-time digital signature, which will be then inserted into the DS field. This digital signature is verified at the BEI-server (in the hospital) prior to accessing the patient record. Four bits allocated for Service code identifies the category of

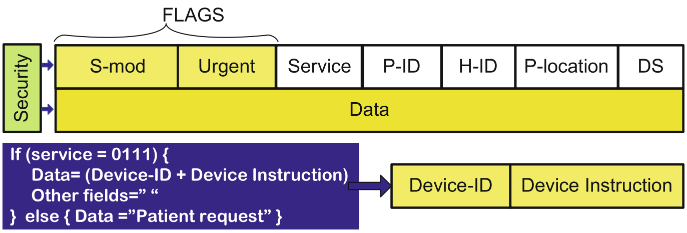


Fig. 3 The data packet format for the remote data exchange of BEI

service provided by BEI. With the current version of BEI, up to 16 different services can be defined. The Service field provides a means to accurately describe medical, surgical, and/or diagnostic services needed for processing patient claims and for medical review. Service field 0111 is used to switch patient agent to insecure mode. In this mode operations of environmental devices for the self-management are performed, including the management of patient bed. This mode also uses four bits for controlling 16 environmental devices. That is, the service code 0111 has 16 other sub-instructions each dedicated to the control of a specific home device or helper functions, which provide a total of self-management tasks.

5 Conclusions

We have designed and tested a significantly inexpensive prototype system applied to real-time monitoring and self-management of patients with verbal communication inabilities. This article gave a brief description about the structure, specifications, functions, and formats of the system, which can also guide the development of similar systems with more complicated technical requirements. The uniqueness of the proposed system lies in the intelligence of its configurable adaption to a variety of patient cares. Due to extensive resource usage of healthcare facilities with the in-hospital patient treatment and monitoring, long term patient cares should be moved more and more to on-site treatment with enhanced self-care capabilities.

It was observed that the application of BEI has some limitations. Since the system requires patient cooperation by not only activated EEG/EMG signals also by the predefined gestures, patients with disorder of consciousness and higher brain dysfunction can hardly benefit from this system. Hence, adequacy of the system must be considered with care for patients of neurologic diseases. It is possible that system and patient training regarding neurologic patients may require longer times to gain maximum advantage of the system.

References

1. Korhonen I, Parkka J, Van Gils M (2003) Health monitoring in the home of the future. IEEE Eng Med Biol Mag 22(3):66–73
2. Komnakos D, Vouyioukas D, Maglogiannis I, Skianis C (2011) Cooperative mobile high-speed and personal area networks for the provision of pervasive e-health services. In: IEEE international conference on communications (ICC), pp 1–5
3. Lee R-G, Chen K-C, Hsiao C-C, Tseng C-L (2007) A mobile care system with alert mechanism. IEEE Tran Inf Technol Biomed 11(5):507–517
4. Hong Y, Patrick T, Gillis R (2008) Protection of patient’s privacy and data security in e-health services. In: International conference on biomedical engineering and informatics BMEI 2008, vol 1, pp 643–647
5. Sulaiman R, Sharma D (2011) Enhancing security in e-health services using agent. In: International conference on electrical engineering and informatics (ICEEI), pp 1–6

6. Elkhodr M, Shahrestani S, Cheung H (2011) Enhancing the security of mobile health monitoring systems through trust negotiations. In: IEEE 36th conference on local computer networks (LCN), pp 754–757
7. International Organization for Standardization (2009) ISO/TS 25237:2008: Health informatics: pseudonymization. International Organization for Standardization, Geneva
8. El Emam K, Jonker E, Arbuckle L, Malin B (2011) A systematic review of re-identification attacks on health data. PLoS One 6(12):e28071
9. El Emam K, Dankar FK, Issa R, Jonker E, Amyot D, Cogo E et al (2009) A globally optimal k-anonymity method for the de-identification of health data. J Am Med Inform Assoc 16(5):670–682
10. Samarati P (2001) Protecting respondents' identities in microdata release. IEEE Trans Knowl Data Eng 13(6):1010–1027

Coverage and Connectivity Guaranteed Deterministic Deployment Pattern for WSN

R. Ramalakshmi and S. Radhakrishnan

Abstract In Wireless Sensor Network, node placement may be random or regular. In deterministic deployment, sensors are placed by hand at selected spots prior to network operation in wireless sensor network. It is more popular today when the deployment area is physically accessible and its success depends on the optimality of the deployment pattern. In this paper, we propose a new ploygon based deployment pattern for optimal placement of sensors in WSN with guaranteed coverage and connectivity. The coverage of the proposed deployment is compared with existing deployments like square, stripe, triangles and shown positive results. The connectivity is measured as the number of active nodes in the dominating set and it is also good when we compare the connectivity with a topology construction algorithms A3 in [7].

1 Introduction

The main goals of a wireless sensor network are to monitor for events with complete coverage and connectivity of the network. Typical applications of sensor network include target tracking, urban monitoring, soil monitoring etc. Each sensor node has a sensing radius within which it can sense data, and a communication radius within which it can communicate with another node. Each of these nodes will collect raw data from the environment, do local processing, possibly communicate with each other in a multihop fashion to sink.

The deployment of sensor nodes in large sensing fields need efficient topology control. One important criterion for being able to deploy an efficient sensor network is to find optimal node placement strategies. Random deployments are used when

R. Ramalakshmi (✉) · S. Radhakrishnan
Kalasalingam University, Anand Nagar, Krishnankoil, India
e-mail: rama@klu.ac.in

S. Radhakrishnan
e-mail: srk@klu.ac.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_34,

the deployment area is physically inaccessible. In random deployment, the sensors are randomly scattered. It is difficult to find a random deployment strategy that minimizes cost, reduces computation and communication, is resilient to node failures, and provides a high degree of area coverage. In regular deployment, sensors are placed at selected spots prior to network operation. This type of deployment have benefits like cost savings, minimizing message overhead and easy network management.

Thus deployment is an important issue in wireless sensor networks and it also affects the network performance. The large number of sensor node placement leads to scalability and reliability problems due to unpredictable nature of deployment conditions.

2 Related Works

An overview of recently proposed deployment schemes for coverage and connectivity are surveyed in [4], and the performance of existing methodologies is discussed. They have also pointed out some crucial open issues in sensor deployment.

A study on deployment patterns to achieve full coverage and k-connectivity is presented under different ratios of the sensor communication range to the sensing range for homogeneous wireless sensor networks in [8]. They propose new patterns for 3- and 5-connectivity. In [2], an unregular equilateral hexagon based deployment pattern is proposed with sensing radius and the communication radius having a different proportion.

In [9], issues of maintaining sensing coverage and connectivity by keeping a minimum number of sensor nodes are addressed. They prove that if the radio range is at least twice the sensing range, complete coverage of a convex area implies connectivity among the working set of nodes. The authors in [1], present different optimal lattice-based grids for arbitrary sensing and communication radius based on squares, triangles, rhombusus, hexagons and strips which provide a minimal connected topology with total coverage.

In [7], authors have proposed different algorithms for reduced toplogy construction using connected dominating set without the need of loacalization information. They compared their work with optimal theoretical bound for connected coverage in sparse and dense networks. They have also compared their work in terms of number of active nodes and the ratio of coverage as main performance metrics.

The toplogy construction algorithms focus to build a reduced topology while preserving important network characteristics like connectivity and coverage. Topology construction and maintenance algorithms are presented in [3, 5].

The rest of this paper is organized as follows: The proposed deployment pattern is explained in Sect. 3. In Sect. 3, the coverage and connectivity of the pattern are verified with coverage algorithms in [7] using a tool Atarraya proposed in [6].

3 Polygon Based Optimal Deployment Pattern

Topology Control is one of the important techniques utilized to reduce energy consumption in wireless sensor networks. The topology construction algorithms aim to reduce the topology for communication. We can model a wireless sensor network as Unit Disk Graph (UDG) where each node is represented as a disk. We have considered a homogeneous sensor network, where all the sensor nodes are identical in terms of sensing and communication capabilities. The sensing and communication radius of a sensor node are represented by r_c and r_s respectively. Each sensor is capable of monitoring the events within distance r_s and it is communicating with other sensor within distance r_c.

The coverage property ensures that any point on the area of interest can be monitored and the connectivity ensures that the network of active sensors are always connected. Coverage implies connectivity when radio range is larger than twice of detecting range. When $r_c = r_s$, the sensing and communication radius is given in Fig. 1a. In an ideal case

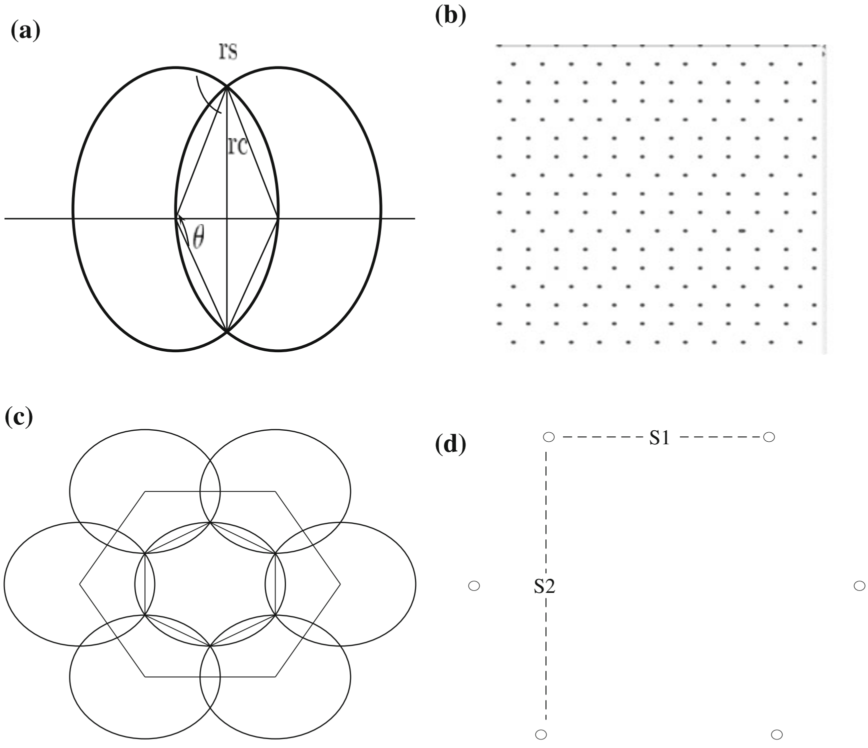


Fig. 1 Topology control. **a** Communication and sensing range. **b** Sensor deployment. **c** Polygon pattern. **d** Coverage area

$$\cos\frac{\theta}{2} = \frac{r_c}{2r_s}$$

From the literature, we find that the given or $\frac{r_c}{r_s} = \sqrt{2}$ will generate the square patterns. So we take initially $\frac{r_c}{r_s} = \sqrt{2}$ and increase this ratio upto $\sqrt{3}$ since it will generate the hexagonal like pattern as in Fig. 1c. The proposed polygon like pattern consists of 7 sensors to cover a particular area. The placement of sensors in the deployment area is shown in Fig. 1b.

$$\sqrt{2} < \frac{r_c}{r_s} < \sqrt{3}$$
$$\frac{\sqrt{2}}{2} < \frac{r_c}{2r_s} < \frac{\sqrt{3}}{2}$$
$$\frac{1}{\sqrt{2}} < \frac{r_c}{2r_s} < \frac{\sqrt{3}}{2}$$
$$\frac{1}{\sqrt{2}} < cos\frac{\theta}{2} < \frac{\sqrt{3}}{2}$$

From the Fig. 1d, we can find S1 and S2 as

$$S1 = 2r_s cos\frac{\theta}{2} * 2sin\frac{\theta}{2}$$
$$S2 = 2r_s cos\frac{\theta}{2} * 2cos\frac{\theta}{2}$$

We can see that each sensor will cover half of the rectangular area appropriately at maximum. So the covered area can be calculated as

$$A = \frac{S1 * S2}{2}$$
$$A = \frac{(2r_s cos\frac{\theta}{2} * 2sin\frac{\theta}{2}) * (2r_s cos\frac{\theta}{2} * 2cos\frac{\theta}{2})}{2}$$
$$A = 2sin\theta(1 + cos\theta)r_s^2$$
$$A = (2sin\theta + sin2\theta)r_s^2$$

If we are deploying in the area R, then the number of polygons required (N) is

$$N = \frac{R}{A} = \frac{R}{(2sin\theta + sin2\theta)r_s^2}$$

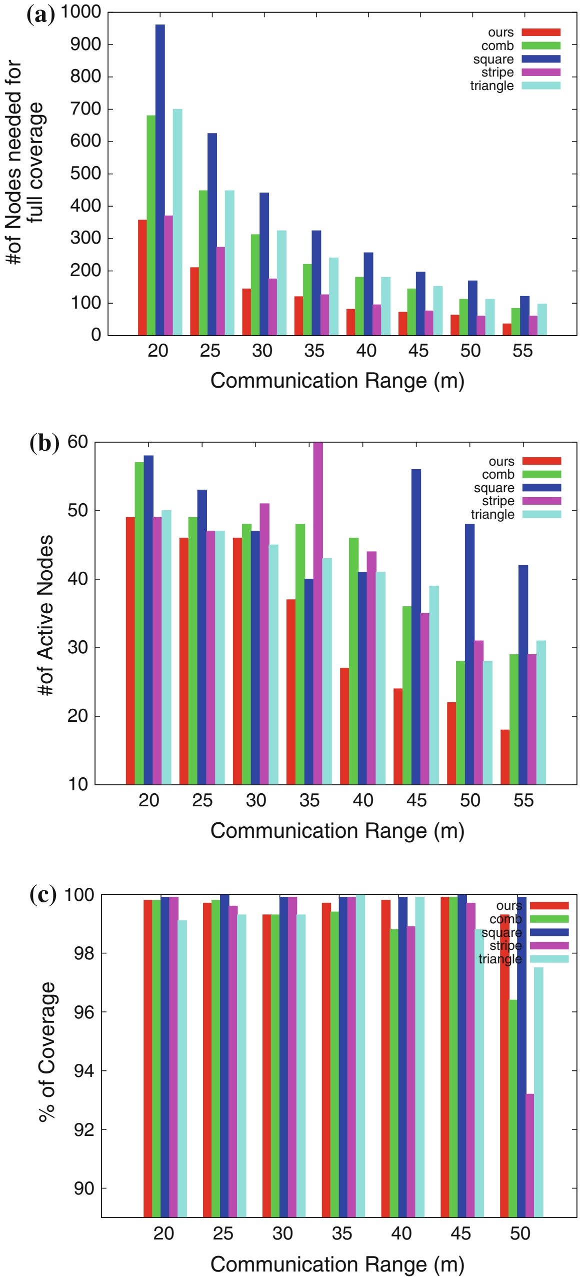


Fig. 2 Connectivity and coverage. **a** Total nodes for full coverage. **b** CDS size. **c** Coverage area

4 Simulation Study

We used the Atarraya tool in [6]. We have used the terrain size as 600 × 600. In the simulation, we used sensors with sensing range $r_s = 20$ m and the communication range r_c various from 20 to 55 m. The number of nodes needed for covering this entire area is shown in Fig. 2a. From the results it is clear that, the proposed ploygon based deployment gives minimum number of nodes compared to other regular deployment patters. The number of sensors needed for full coverage decreases when the ratio $\frac{r_c}{r_s}$ increases.

A toplogy construction algorithm for selecting active nodes called dominating nodes is proposed in [7]. We have verified the connectivity of the sensor network by running A3 algorithm in Atarraya. The result of number nodes in the active set generated by A3 is presented in Fig. 2b. The size is much smaller than other deployments. The results shows that minimum number of nodes are active with connectivity and full coverage.

The coverage region is presented in Fig. 2c. The result shows that the proposed deployment is giving more than 99 % of coverage like other regular deployment patterns.

5 Conclusion

In this paper, we have proposed a deterministic deployment pattern for wireless sensor networks. The proposed work is compared with other regular deployment patters like square, strip, triangle and honeycomb in terms of coverage and connectivity. The results show that, the proposed polygon based pattern achieves 99 % coverage. The connectivity is measured by taking the number of active nodes in the dominating set that participate in forwarding of data to the base station. The active nodes based connectivity is also good. We plan to extend this work to verify the connectivity for k-coverage.

References

1. Bai H, Xuan D, Yun Z, Lai T, Jia W (2008) Complete optimal deployment patterns for full-coverage and k-connectivity ($k \leq 6$) wireless sensor networks. In: Proceedings of the 9th ACM international symposium on mobile ad hoc networking and computing.
2. Junguo Z, Feng Z (2012) Study on optimal regular deployment patterns of wireless sensor network. Res J Appl Sci Eng Technol 4:2300–2303
3. Labrador MA, Wightman PM (2009) Topology control in wireless sensor networks. Springer, Heidelberg
4. Liu L, Xia F, Wang Z, Chen J, Sun Y (2005) Deployment issues in wireless sensor networks. Lecture notes in computer science, pp 239–248.
5. Santhi P (2005) Topology control in ad hoc and sensor networks. Wiley, New Jersey

6. Wightman P, Labrador M (2009) Atarraya-a simulation tool to teach and research topology control algorithms for wireless sensor networks. In: Proceedings of the 2nd international ICST conference on simulation tools and techniques.
7. Wightman PM, Labrador MA (2011) A family of simple distributed minimum connected dominating set-based topology construction algorithms. J Netw Comput Appl 1997–2010.
8. Yun Z, Bai X, Xuan D, Lai TH, Weijia J (2010) Optimal deployment patterns for full coverage and k-connectivity ($k \leq 6$) wireless sensor networks. IEEE/ACM Trans Netw 18:934–947
9. Zhang H, Hou J (2005) Maintaining sensing coverage and connectivity in large sensor networks. Ad Hoc Sens Wirel Netw 1:89–123

Hybrid Deployment Schemes for Wireless Sensor Networks

G. Sanjiv Rao and V. Vallikumari

Abstract The Wireless Sensor Networks are used in many applications in military, ecological, health related areas, for example, rapid deployment, self-organization, and fault tolerance characteristics of sensor networks make them a very promising technique for military command, control, communications and the targeting systems. Sensor networks are expected to pay an essential role in the upcoming age of pervasive computing. These networks are deployed in harsh and inaccessible environments with the purpose of monitoring their respective surroundings, and generating observed readings. Deployment of nodes in Wireless Sensor Networks (WSNs) is a basic issue to be addressed as it can influence the performance metrics of WSNs connectivity, resilience and scalability requirements. Many deployment schemes have been proposed for wireless sensor networks. We survey six deployment models random, rectangular-grid, square-grid, triangular-grid, hexagonal-grid, grid-group, and proposed two novel schemes for deployment, combinational and hybrid deployment schemes, to show which deployment scheme can be used to increase network connectivity, and scalability requirements with respect to some factor. We present the analytical and simulation-based results of the WSN made up of mica2 motes using the deployment knowledge to motivate the use of these emerging paradigms. WSNs have been simulated with Network Simulator 2.34 for node configuration, sink node configuration, topology creation, and node configured with sensing, temperature and energy capabilities by using mannasim.

Keywords Wireless sensor network · Deployment knowledge · Connectivity · Scalability

G. Sanjiv Rao (✉)
Department of IT, Sri Sai Aditya Institute of Science and Technology, Kakinada, India
e-mail: sanjiv_gsr@yahoo.com

V. Vallikumari
Department of CS & SE, College of Engineering, Andhra University, Visakhapatnam, India
e-mail: vallikumari@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_35,

1 Introduction

Wireless sensor network (WSN) is a collection of tiny, low powered battery devices called sensor nodes. The nodes communicate with each other wirelessly and the resulting network is usually used for monitoring an environment by gathering local data such as temperature, light or motion [1]. These tiny sensor nodes, consisting of sensing, processing, and communication components, make it possible to create wireless sensor networks (WSNs), which provides a significant improvement over traditional wired sensor networks [2]. The sensor nodes in a WSNs are deployed over a targeted area to sense and gather various types of data that includes temperature, humidity, intrusion detection, vehicular motion and so on [3]. A sensor node in WSNs is also known as a mote, which is capable of performing some task such as processing, information gathering and communicating with other connected nodes in the network. The transmission between the sensors is done by short range radio communications. The base station is assumed to be computationally well-equipped whereas the sensor nodes are resource- starved. The sensor nodes are usually scattered in a sensor field (i.e., deployment area or target field) as shown in Fig. 1 each of these scattered sensor nodes has the capabilities to collect data and route data back to the base station. Data are routed back to the base station by a multi-hop infrastructureless architecture through sensor nodes. The base station may communicate with the task manager node via internet or satellite. An ad-hoc network is a group of mobile, wireless hosts which co-operatively and spontaneously form a network independently of any fixed infrastructure or centralized administration. In particular, an ad-hoc network has no base stations, host, also called node, communicates directly with nodes within its wireless range and indirectly with all other destinations using a multi-hop route through other nodes in the network [3].

2 Deployment Schemes for WSNs

In WSNs, the major challenge is the deployment of the nodes in the deployment region to satisfy continuous sensing with extended network lifetime while maintaining uniform coverage. Various architectures and node deployment strategies have been developed for wireless sensor networks, depending upon the requirement of its application [4]. In our study, we illustrate various node deployment strategies for wireless sensor networks random, rectangular-grid, square-grid, triangular-grid, hexagonal-grid, grid-group, combinational and hybrid deployment schemes. We analyze various performance metrics connectivity and scalability with respect to each of these deployment schemes.

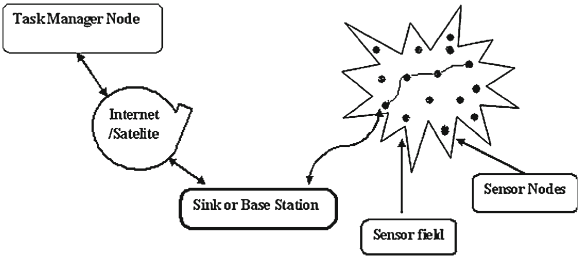


Fig. 1 Sensor nodes scattered in a target field

3 Our Contribution

In this paper WSN has been simulated with Network Simulator 2.34 for node configuration, sink node configuration, topology creation, and sensor nodes have been deployed using random, rectangular-grid, square-grid, triangular-grid, hexagonal-grid, grid-group, combinational deployment and hybrid deployment schemes, and sensor node has been configured with sensing, temperature, energy capabilities by using Mannasim.

3.1 Simulation with NS-2.34 and Mannasim

NS-2.34 is a discrete event network simulator that has developed in 1989 as a variant of the REAL network simulator. Initially intended for wired networks, the Monarch Group at CMU has extended NS-2.34 to support wireless networking. But to simulate WSN in NS-2.34, it needs to have additional module to represent the protocols specific to WSN.

MANNASIM is a framework for WSN simulation based on NS-2.34. It extends NS-2.34 by introducing new modules for design, development and analysis of different WSN applications. The goal of MANNASIM is to develop a detailed simulation framework, which can accurately model different sensor nodes and applications while providing a versatile test bed for algorithms and protocols.

- Example of Antenna settings
  - Antenna/OmniAntenna set X_ 0
  - Antenna/OmniAntenna set Y_ 0
  - Antenna/OmniAntenna set Z_ 1.5
  - Antenna/OmniAntenna set Gt_ 1.0
  - Antenna/Omni Antenna set Gr_ 1.0
- Example of sensing capabilities
  - Node/MobileNode/SensorNode set sensingPower_ 0.015

- Node/MobileNode/SensorNode set processing Power_ 0.024
- Node/MobileNode/SensorNode set instructionsPerSecond_ 8000000
- Setting up mica 2 mote with $antenna and range = $range
- Phy/WirelessPhy set Pt_ 0.281838
- Phy/WirelessPhy set freq_ 2.4e09
- Phy/WirelessPhy set L_ 1.0
- Phy/WirelessPhy set lambda_ 0.125
- Phy/WirelessPhy set RXThresh_ [TwoRay 0.281838 [$antenna set Gt_] [$antenna set Gr_] 0.8 0.8 1.0 $range 0.125]

3.2 Simulation of Random Deployment

As explained in [5], random deployment refers to the situation in which sensor nodes are uniformly and independently distributed across the monitored field. When practical application scenarios are considered, random deployment is a feasible and practical method, and sometimes it is the only feasible strategy [6]. Random approach for node deployment is deeply discussed in [10], which has considered as one of the competitors. In this deployment, as shown in Fig. 2a we visualize a simulated sensor network with 20 nodes deployed with random deployment scheme, each of the sensors has equal probability of being placed at any point inside a given target field.

3.3 Simulation of Grid Deployment

In [6], it has state that grid deployment is an attractive approach for moderate to large-scale coverage-oriented deployment due to its simplicity and scalability. Based on these considerations, we focus on popular grid layouts such as rectangular-grid, square-grid, triangular-grid, and hexagonal-grid. For the coverage performance, grid-based deployment schemes are considered as a good deployment schemes in WSNs.

3.3.1 Simulation of Rectangular Grid Deployment

In Fig. 2b shows the visualization of a simulated sensor network with 20 nodes deployed with rectangular grid deployment scheme. The desired distance between consecutive droppings is achieved by controlling the time intervals. However, due to placement errors, this ideal deployment is not realistic.

3.3.2 Simulation of Square Deployment

As described in [6], we investigate a square grid because of its natural placement strategy over a unit square. In Fig. 2c we present the simulation results of a square

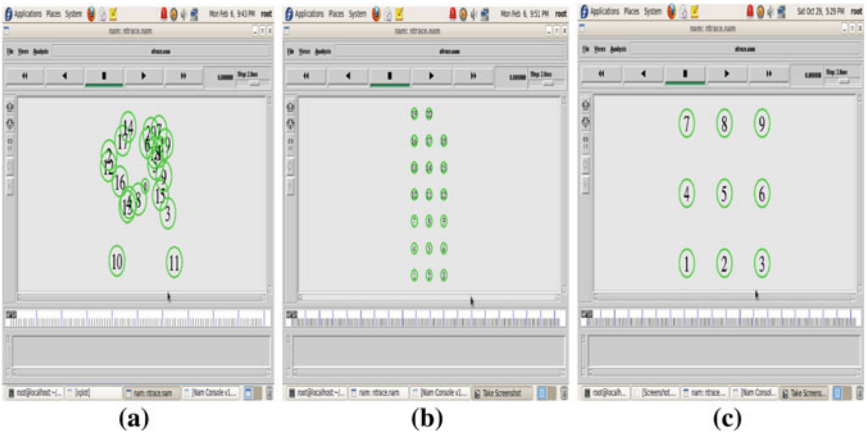


Fig. 2 **a** Simulation of random deployment; **b** simulation of rectangular grid deployment; **c** simulation of square grid deployment

grid deployment of 9 sensor nodes. We calculated the approximate length and area of a unit square by using the formula given in [6].

3.3.3 Simulation of Triangular Grid Deployment

As explained in [7], in this triangular deployment scheme can cover a sensing area fully. It divides the target area into cells. Each cell represents the triangular-shaped sensing area. In Fig. 3a we presented the simulation-based result of triangular deployment scheme. This figure visualizes triangle deployment of 16 nodes among which node 0 represents sink node.

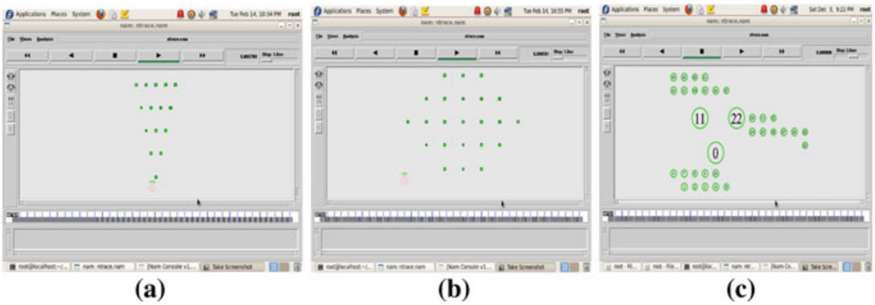


Fig. 3 **a** Simulation of triangular grid; **b** simulation of hexagonal grid; **c** simulation of grid group deployment

3.3.4 Hexagonal Grid Deployment

In this paper, we assume that the wireless sensor networks are made up of mica2 motes and have an omni directional antenna. Therefore, each node has a round shape of communication area. One of the main goals of our work is to improve the covering area of interest to guarantee network connectivity. We assumed a semi-regular tiling which uses triangle and hexagon in the two dimensional plane instead of considering a regular polygon which has the same side lengths and interior angle as described in [8]. As specified in [9], a tiling or tessellation of the plane is a process which uses a collection of polygons called tiles to cover the plane without gaps or overlaps. Figure 3b shows a hexagonal grid deployment simulation, in which a target field is partitioned hexagonal grids. This figure visualizes the simulation of 24 nodes in the target area using hexagonal grid deployment to improve the network connectivity.

3.3.5 Grid-Group Deployment

Grid-group deployment scheme for wireless sensor networks is discussed in [10], when nodes are deployed in a region, due to limited power, all nodes cannot communicate with all other nodes in the network. So we divide the entire region into equal-sized squares or grids as done in Liu and Ning [2003, 2005], and Huang and Medhi [2007]. This scheme has the advantage that all nodes within a particular region can communicate with each other directly and nodes which lie in a different region can communicate via special nodes called agents which have more resources than the general nodes, it is also assumed that it is more difficult to compromise an agent than a sensor node. Whatever the size of the network, the number of agents in a region is always three. This scheme ensures that even if one region is totally disconnected, the regions are not affected.
Figure 3c shows the visualization of a simulated sensor network with 10 nodes deployed with each grid- group deployment scheme, in this figure there are three sink nodes 0, 11, 22 and each sink node consists 10 nodes deployed in a grid manner.

3.4 Combinational Deployment

As we have specified in the previous sections, in practice, random deployment experiences excess redundancy to overcome uncertainty and it could be very expensive, and grid deployment is often infeasible to guarantee exact placement due to various errors, including misalignments and random misplacement. Here, we had proposed one novel deployment scheme which is a combination of both random and grid deployments that could provide the advantages of random as well as grid deployment schemes. We had implemented this combinational deployment by deploying some nodes by using grid model to provide coverage performance and then, some nodes were deployed randomly to overcome the placement errors.

We had provided the visualization for the combinational deployment model of sensor nodes in the target field in Fig. 4a in which the visualization consists of deployment of 20 nodes. Among the 20 nodes, first 10 nodes have been deployed in grid fashion. The desired distance between consecutive droppings is achieved by controlling the time intervals. However, to avoid placement errors, the remaining 10 nodes have been deployed randomly. In Fig. 4b visualizes simulation of combinational deployment scheme in which nodes have been configured with sensing, propagation, processing, and data generator capabilities.

3.5 Hybrid Deployment

In the previous sections we survey possible deployment schemes for WSNs each of these schemes has their own advantages and disadvantages. In order to gain the beneficial effects of all these schemes in the area of WSNs, we had proposed another naïve node deployment scheme, the hybrid deployment scheme which is a combination of all the previously described deployment schemes. In hybrid deployment, the target field consists of nodes deployed using various schemes such as random, rectangular-grid, square-grid, hexagonal-grid, grid-group, and combinational deployments to improve network connectivity, scalability. We study performance of WSNs based on node distribution. Compared with existing schemes, hybrid node deployment scheme requires a shorter transmission range and achieves a higher connectivity and scalability. In Fig. 5a shows the simulation result of a hybrid deployment which is a combination of all other schemes specified in the previous sections. We simulated the WSN consists of 44 nodes among which some nodes were deployed using grid model, some nodes were deployed group deployment and some node were deployed using random model etc., In this simulation nodes 0, 11, 22, 33, and 44 are the sink nodes. Figure 5b shows that WSNs has been simulated and sensor nodes have been

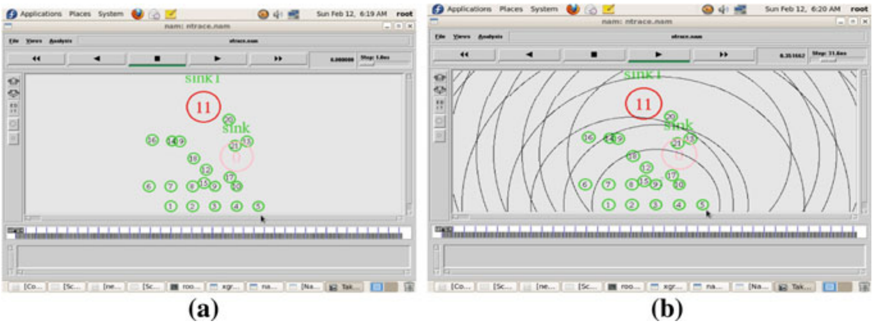


Fig. 4 **a** Simulation of combinational deployment; **b** Sensing capabilities for combinational deployment

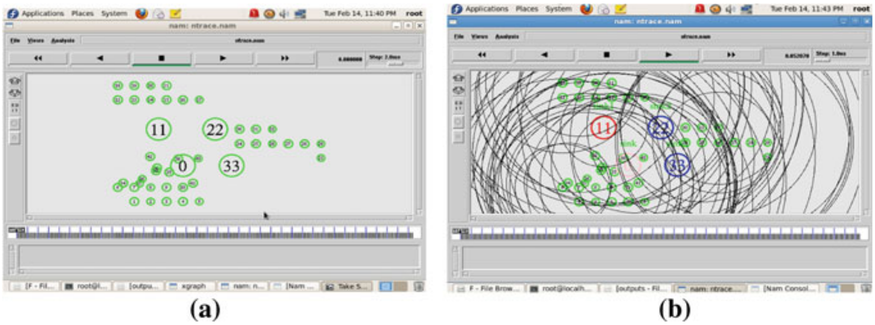


Fig. 5 **a** Simulation of hybrid deployment; **b** sensing capabilities for hybrid deployment

Table 1 Comparison of the different schemes with respect to number of nodes, connectivity, scalability

Deployments schemes	Sensor mote	Number of nodes	Connectivity	Scalability
Random	Mica 2	20 nodes	Very low	Scalable
Rectangular	Mica 2	20 nodes	Low	Not scalable
Square gird	Mica 2	9 nodes	Low	Not scalable
Triangular grid	Mica 2	16 nodes	Medium	Not scalable
Hexagonal grid	Mica 2	24 nodes	High	Not scalable
Grid-group	Mica 2	33 nodes	High	Not scalable
Combinational	Mica 2	22 nodes	High	Scalable
Hybrid	Mica 2	44 nodes	Very high	Scalable

deployed by hybrid deployment scheme, hence all the nodes can communicate each other in between the hybrid deployment which consists of the sensing capabilities, propagation, processing capability, data generator, setting have been configured.

3.6 Comparison with Performance Metrics Connectivity, Scalability

Thus, so far we have only considered various node deployment schemes for wireless sensor networks in this we describe the comparative study of these node deployment schemes (Table 1).

4 Conclusions

In this paper, we focus on various node deployment schemes for wireless sensor networks, and our study is based on a network model in which homogeneous nodes distributed in a target area. We also simulated WSN with NS-2.34 and Mannasim and sensor nodes have been deployed by using random, rectangular-grid, square-grid, triangular-grid, hexagonal-grid, grid-group, combinational and hybrid deployment schemes and configured the sensing parameters and also presented the comparative study of those schemes with respect to various performance metrics connectivity and scalability. It can be extended for WSNs security requirements, for secure data transmission.

References

1. Kendall M, Martin K (2011) On the role of expander graphs in key predistribution schemes for wireless sensor networks. WEWoRC 2011:62–82 (Information Security Group, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, UK)
2. Yong W, Garhan A, Byrav R (2006) A survey of security issues in wireless sensor networks. IEEE Commun Surv Tutor (2nd Quarter) 8:2–23 (University of Nebraska-Lincoln)
3. Akyildiz IF, Su W, Sankarasubramaniam Y (2002) A survey on sensor networks. IEEE Commun Mag 40:102–114
4. Messai M-L, Aliouat M, Seba H (2010) Tree based protocol for key management in wireless sensor networks. EURASIP J Wireless Commun Netw (research Article) 2010:910695
5. Fan G, Wang R, Huang H, Sun L, Sha C (2010) Coverage-guaranteed sensor node deployment strategies for wireless sensor networks. Sensors 10:2064–2087
6. Monica, Sharma AK (2010) Comparative study of energy consumption for wireless networks based on random and grid deployment strategies. Int J Comput Appl (0975–8887), 6(1):28–35
7. Yu Z, Guan Y (2008) A key management scheme using deployment knowledge for wireless sensor networks. IEEE Trans Parallel Distrib Syst 19(10):1411–1425
8. Wu C-H, Chung Y-C (2010) A tiling-based approach for directional sensor network deployment. In: IEEE SENSORS 2010 conference
9. Grunbaum B, Shephard GC (1977) Tilings by regular polygons. Math Mag 50(5):227–247
10. Liu D, Ning P, Du W (2008) Group-based key predistribution for wireless sensor networks. ACM Trans Sens Netw 4(2):Article 11

Adaptive Optimal Distributed Power Allocation for Enhanced Cognitive Radio Network in the Presence of Channel Uncertainties

Hao Xu and S. Jagannathan

Abstract In this paper, novel enhanced Cognitive Radio Network (CRN) is considered by using power control where secondary users (SUs) are allowed to use wireless resources of the primary users (PUs) when PUs are deactivated, but also allow SUs to coexist with PUs while PUs are activated by managing interference caused from SUs to PUs. Therefore, a novel adaptive optimal distributed power allocation (AODPA) scheme is proposed by incorporating the effect of channel uncertainties for enhanced CRN in the presence of wireless channel uncertainties. In order to mitigate the attenuation of SIR's due to channel uncertainties the proposed novel AODPA allows the SIR's of both PUs' and SUs' to converge to a desired target SIR while minimizing the energy consumption. Simulation results illustrate that this novel AODPA scheme can converge much faster and cost less energy than others by adapting to the channel variations optimally.

Keywords Adaptive optimal distributed power allocation (AODPA) · Signal-to-interference ratio (SIR) · Channel uncertainties · Cognitive radio network

1 Introduction

Cognitive Radio Network (CRN) [1] is a promising wireless network since CRN can improve the wireless resource (e.g. spectrum, power etc.) usage efficient significantly by implementing more flexible wireless resource allocation policy. In [2], a novel secondary spectrum usage scheme (i.e. opportunistic spectrum access) is introduced.

H. Xu (✉) · S. Jagannathan
Department of Electrical and Computer Engineering,
Missouri University of Science and Technology, Rolla, MO, USA
e-mail: hx6h7@mst.edu

S. jagannathan
e-mail: sarangap@mst.edu

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_36,

The SUs in SRN can access the spectrum allocated to PUs originally while the spectrum is not used by any PU.

An efficient transmission power allocation cannot only improves the network performance (e.g. spectrum efficient, network throughput etc.), but also guarantees the Quality-of-Service (QoS) of PUs. A traditional scheme to protect transmission of PUs is introduced in [3] by imposing a power constraint less than a prescribed threshold referred to as interference temperature constraint in order to contain the interference caused by SUs to each PU. To relax the centralized power allocation (CPA) for balancing the signal-to-interference ratios (SIR) of all PUs and SUs, distributed power allocation (DPA) is proposed in [4] for enhanced CRN since information from other PUs and SUs is not needed. However, wireless channel uncertainties are not considered in these works [1–4].

For incorporating channel uncertainties into DPA, Jagannathan and Zawodniok [5] proposed a novel DPA algorithm to maintain a target SIR for each wireless receiver in cellular network under channel uncertainties. In [5], an adaptive estimator (AE) is derived to estimate slowly time varying SIR model which can be changed with varying power and channel uncertainties, and then adaptive DPA is proposed to force actual SIR of each wireless user converge to target SIR. Motivated from this idea, channel uncertainties have also been included into the developed novel adaptive optimal DPA scheme.

In this paper, a novel adaptive optimal distributed power allocation (AODPA) for PUs and SUs in enhanced CRN with channel uncertainties is proposed by using adaptive dynamic programming (ADP). Based on the special property of enhanced CRN (i.e. introduced SUs can use PU's wireless resource when PUs are deactivated, also SUs are allowed to coexist with PUs while PUs are activated by managing interference caused from SUs to PUs properly), AODPA can be developed under two cases: Case 1 PUs are deactivated while in Case 2 PUs are activated. In Case 1, since PUs are deactivated and SUs would dominant CRN, proposed AODPA has to force the SIRs of SUs to converge to a higher target value in order to increase the CRN utility. However, in Case 2, since PUs are activated, proposed AODPA has to not only force the SIRs of the SUs to converge to a low target value to guarantee QoS for PUs, but also increase network utility by allocating the transmission power properly for both PUs and SUs. Meanwhile, proposed AODPA algorithm being highly distributive in nature does not require any inter-link communication, centralized computation, and reciprocity assumption.

2 Background

As shown in Fig. 1, the general enhanced Cognitive Radio Network (CRN) can be classified into two types of sub-networks: Primary Radio Network (PRN) and Secondary Radio Network (SRN) which include all PUs and SUs in enhanced CRN respectively. In order to improving network utilities (e.g. spectrum efficiency etc.), SUs are introduced in enhanced CRN to share the wireless resource

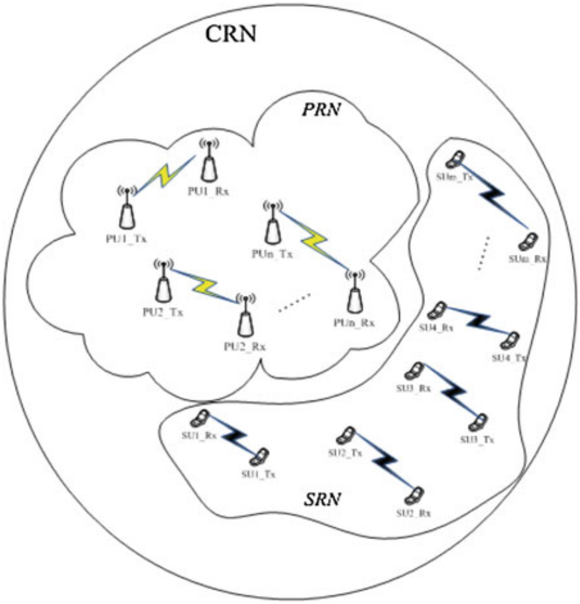


Fig. 1 Enhanced cognitive radio network

(e.g. spectrum etc.) with PUs which usually exclusive network resource in other existing wireless networks (e.g. WLAN, WiMAX, etc.)

Due to special property of enhanced CRN, traditional wireless network protocol (e.g. resource allocation, scheduling etc.) might not be suitable for CRN. Therefore, novel protocol is extremely needed to be developed for enhanced CRN. Using the enhanced CRN property, novel protocol has to be separated into two cases: Case 1 PUs are deactivated; Case 2 PUs are activated. In Case 1, since SUs dominant the enhanced CRN, enhanced CRN network protocol has to improve the SRN performance as much as possible. However, in Case 2, enhanced CRN network protocol has to not only guarantee QoS of PUs, but also increase CRN network utility by allocating resource to both PUs and SUs properly.

3 Proposed Adaptive Optimal Distributed Power Allocation (AODPA) Scheme

3.1 *Dynamic SIR Model for PUs & SUs with Unknown Uncertainties*

In previous power allocation schemes [3, 4], only path loss uncertainty is considered normally. Without considering the mobility of PUs and SUs, the mutual interference $I(t)$ is held constant which is actually inaccurate in practical cognitive radio network.

Therefore, in this paper, more uncertainties factors included path loss, shadowing and Rayleigh fading are considered together and both channel gain h and the mutual interference $I(t)$ are assumed to be slowly time-varying. According to the SIRs, $R_l^{PU}(t) R_m^{SU}(t)$, at the receiver of lth PU and mth SU at the time instant t can be calculated respectively as:

$$\begin{aligned} R_l^{PU}(t) &= \frac{h_{ll}(t) P_l^{PU}(t)}{I_l^{PU}(t)} = \frac{h_{ll}(t) P_l^{PU}(t)}{\sum\limits_{l \neq j \in \{PUs\}} h_{lj}(t) P_j^{PU}(t) + \sum\limits_{i \in \{SUs\}} h_{lj}(t) P_i^{SU}(t)} \\ R_m^{SU}(t) &= \frac{h_{mm}(t) P_m^{SU}(t)}{I_m^{SU}(t)} = \frac{h_{mm} P_m^{SU}(t)}{\sum\limits_{j \in \{PUs\}} h_{mj}(t) P_j^{PU}(t) + \sum\limits_{m \neq i \in \{SUs\}} h_{mi}(t) P_i^{SU}(t)} \end{aligned} \tag{1}$$

where $I_l^{PU}(t)$, $I_m^{SU}(t)$ is the mutual interference for lth PU and mth SU, $P_l^{PU}(t)$, P_m^{SU} are the transmitter power of lth PU and mth SU,and $\{PUs\}$, $\{SUs\}$ are the sets of PUs and SUs respectively.

Differentiating Eq. (1) on both sides and using Euler's formula, differential equation of Eq. (1) can be transformed to discrete-time domain and expressed as

$$\begin{aligned} R_{l,k+1}^{PU} &= \phi_{l,k}^{PU} R_{l,k}^{PU} + \rho_{l,k}^{PU} \upsilon_{l,k}^{PU} \\ R_{m,k+1}^{SU} &= \phi_{m,k}^{SU} R_{m,k}^{SU} + \rho_{m,k}^{SU} \upsilon_{m,k}^{SU} \end{aligned} \tag{2}$$

where the variables $\{\phi_{l,k}^{PU}, \rho_{l,k}^{PU}, \upsilon_{l,k}^{PU}\}$ and $\{\phi_{m,k}^{SU}, \rho_{m,k}^{SU}, \upsilon_{m,k}^{SU}\}$ for lth PU and mth SU are defined respectively as $\phi_{l,k}^{PU} = \frac{h_{ll,k+1} - h_{ll,k}}{h_{ll,k}} - \frac{1}{I_{l,k}^{PU}} \Big[\sum\limits_{j \neq l \in \{PUs\}} [(h_{lj,k+1} - h_{lj,k}) P_{j,k}^{PU} + (P_{j,k+1}^{PU} - P_{j,k}^{PU}) \times h_{lj,k}] + \sum\limits_{i \in \{SUs\}} [(h_{li,k+1} - h_{li,k}) P_{i,k}^{SU} + (P_{i,k+1}^{SU} - P_{i,k}^{SU}) h_{li,k}] \Big]$; $\phi_{m,k}^{SU} = \frac{h_{mm,k+1} - h_{mm,k}}{h_{mm,k}} - \frac{1}{I_{m,k}^{SU}} \Big[\sum\limits_{i \neq m \in \{SUs\}} [(h_{mi,k+1} - h_{mi,k}) P_{i,k}^{SU} + (P_{i,k+1}^{SU} - P_{i,k}^{SU}) h_{mi,k}] + \sum\limits_{j \in \{PUs\}} [(h_{mj,k+1} - h_{mj,k}) P_{j,k}^{PU} + (P_{j,k+1}^{PU} - P_{j,k}^{PU}) h_{mj,k}] \Big]$; $\rho_{l,k}^{PU} = h_{ll,k}$; $\rho_{m,k}^{SU} = h_{mm,k}$; $\upsilon_{l,k}^{PU} = \frac{P_{l,k+1}^{PU}}{I_{l,k}^{PU}}$ and $\upsilon_{m,k}^{SU} = \frac{P_{m,k+1}^{SU}}{I_{m,k}^{SU}}$.

Using Eq. (2), SIR dynamics of each PU (i.e. $l \in \{PUs\}$) and SU (i.e. $m \in \{SUs\}$) can be obtained without loss of generality. Moreover, it is observed that the SIR dynamics for PUs and SUs is a function of wireless channel variation from time instant k to $k + 1$. However, due to uncertainties, wireless channel variation cannot be known beforehand which causes the DPA scheme development for PUs and SUs more different and challenging, especially for optimal designing.

3.2 *Value Function Setup for Adaptive Optimal DPA in Enhanced CRN*

Case 1: PUs are deactivated

In Case 1, since PUs are deactivated, wireless resource (e.g. spectrum etc.) allocated to PUs will be free. Therefore, proposed AODPA scheme would force SUs' SIRs converge to a high target SIR, γ_H^{SU}, in order to improving the performance of enhanced CRN (e.g. spectrum efficiency, network capacity, etc.). Therefore, the SIR error dynamics for SUs in enhanced CRN can be represented by using Eq. (8) as:

$$\begin{bmatrix} e^{SU,1}_{m,k+1} \\ \gamma_H^{SU} \end{bmatrix} = \begin{bmatrix} \phi^{SU,1}_{m,k} & \phi^{SU,1}_{m,k} - 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e^{SU,1}_{m,k} \\ \gamma_H^{SU} \end{bmatrix} + \begin{bmatrix} \rho^{SU,1}_{m,k} \\ 0 \end{bmatrix} \upsilon^{SU,1}_{m,k} \tag{3}$$

In the other words, $E^{SU,1}_{m,k+1} = A^{SU,1}_{m,k} E^{SU,1}_{m,k} + B^{SU,1}_{m,k} \upsilon^{SU,1}_{m,k}$ with SIR error in Case 1 as $e^{SU,1}_{m,k} = R^{SU,1}_{m,k} - \gamma_H^{SU}$, γ_H^{SU} is the high target SIR for SUs under Case 1, and augmented state $E^{SU,1}_{m,k} = [\, e^{SU,1}_{m,k} \;\; \gamma_H^{SU} \,]^T$. Then, according to Eq. (3), the cost function for mth SU in Case 1 can be defined as $J^{SU,\,1}_{m,\,k} = (E^{SU,\,1}_{m,\,k})^T G^{SU,\,1}_{m,\,k} E^{SU,\,1}_{m,\,k}$ with $G^{SU,\,1}_{m,\,k}$ is the solution of the Riccati equation [6].

Using Bellman equation and cost function definition, we can formulate the following equation by substituting value-function into Bellman equation as

$$\begin{bmatrix} E^{SU,1}_{m,k} \\ \upsilon^{SU,1}_{m,k} \end{bmatrix}^T \Theta^{SU,1}_{m,k} \begin{bmatrix} E^{SU,1}_{m,k} \\ \upsilon^{SU,1}_{m,k} \end{bmatrix} = r(E^{SU,1}_{m,k}, \upsilon^{SU,1}_{m,k}) + J^{SU,1}_{m,k+1} \tag{4}$$

Therefore, slowly time varying $\Theta^{SU,1}_{m,k}$ matrix can be expressed as

$$\begin{aligned} \Theta^{SU,1}_{m,k} &= \begin{bmatrix} \Theta^{EE,SU,1}_{m,k} & \Theta^{E\upsilon,SU,1}_{m,k} \\ \Theta^{\upsilon E,SU,1}_{m,k} & \Theta^{\upsilon\upsilon,SU,1}_{m,k} \end{bmatrix} \\ &= \begin{bmatrix} Q^{SU,1} + (A^{SU,1}_{m,k})^T G^{SU,1}_{m,k+1} A^{SU,1}_{m,k} & (A^{SU,1}_{m,k})^T G^{SU,1}_{m,k+1} B^{SU,1}_{m,k} \\ (B^{SU,1}_{m,k})^T G^{SU,1}_{m,k+1} A^{SU,1}_{m,k} & S^{SU,1} + (B^{SU,1}_{m,k})^T G^{SU,1}_{m,k+1} B^{SU,1}_{m,k} \end{bmatrix} \end{aligned}$$

Next, according to [6], the gain of the optimal power allocation for mth SU under Case 1 can be represented in term of value function parameters, $\Theta^{SU,1}_{m,k}$, as

$$\begin{aligned} K^{SU,1}_{m,k} &= [S^{SU,1} + (B^{SU,1}_{m,k})^T G^{SU,1}_{m,k+1} B^{SU,1}_{m,k}]^{-1} (B^{SU,1}_{m,k})^T G^{SU,1}_{m,k+1} A^{SU,1}_{m,k} \\ &= (\Theta^{\upsilon\upsilon,SU,1}_{m,k})^{-1} \Theta^{\upsilon E,SU,1}_{m,k} \end{aligned} \tag{5}$$

It is important to note that if the parameter vector $\Theta_{m,k}^{SU,1}$ can be estimated then mth SU's SIR error dynamic is not needed to calculate optimal DPA gain.

Case 2: PUs are activated

In Case 2, proposed AODPA scheme should not only force PUs' SIRs converge to a desired target SIR (i.e. γ^{PU}) for maintaining their QoS, but also force SU' SIRs converge to a low target SIR (i.e. γ_L^{SU}) in order to coexist with PUs. Therefore, the SIR error dynamics for lth PU and mth SU can be expressed as

$$\begin{aligned} l\text{th}\quad \text{PU}:\quad E_{l,k+1}^{PU,2} &= A_{l,k}^{PU,2}E_{l,k}^{PU,2} + B_{l,k}^{PU,2}\upsilon_{l,k}^{PU,2} \\ m\text{th}\quad \text{SU}: E_{m,k+1}^{SU,2} &= A_{m,k}^{SU,2}E_{m,k}^{SU,2} + B_{m,k}^{SU,2}\upsilon_{m,k}^{SU,2} \end{aligned} \tag{6}$$

where SIR error in Case 2 for PU and SU as $e_{l,k}^{PU,2} = R_{l,k}^{PU,2} - \gamma^{PU}, e_{m,k}^{SU,2} = R_{m,k}^{SU,2} - \gamma_L^{SU}$, γ^{PU}, γ_L^{SU} is the desired target SIR for PUs and high target SIR for SUs under Case 2, and augmented state $E_{l,k}^{PU,2} = [e_{l,k}^{PU,2}\ \gamma^{PU}]^T$, $E_{m,k}^{SU,2} = [e_{m,k}^{SU,2}\ \gamma_L^{SU}]^T$. Then, according to the same theory and derivation as Case 1, the gain of the optimal power allocation for lth PU and mth SU under Case 2 can be represented in term of value function parameters, $\Theta_{l,k}^{PU,2}$, $\Theta_{m,k}^{SU,2}$, respectively as

$$\begin{aligned} K_{l,k}^{PU,2} &= [S^{PU,2} + (B_{l,k}^{PU,2})^T G_{l,k+1}^{PU,2} B_{l,k}^{PU,2}]^{-1}(B_{l,k}^{PU,2})^T G_{l,k+1}^{PU,2} A_{l,k}^{PU,2} \\ &= (\Theta_{l,k}^{\upsilon\upsilon,PU,2})^{-1}\Theta_{l,k}^{\upsilon E,PU,2} \\ K_{m,k}^{SU,2} &= [S^{SU,2} + (B_{m,k}^{SU,2})^T G_{m,k+1}^{SU,2} B_{m,k}^{SU,2}]^{-1}(B_{m,k}^{SU,2})^T G_{m,k+1}^{SU,2} A_{m,k}^{SU,2} \\ &= (\Theta_{m,k}^{\upsilon\upsilon,SU,2})^{-1}\Theta_{m,k}^{\upsilon E,SU,2} \end{aligned} \tag{7}$$

3.3 Model-Free Online Tuning Adaptive Estimator for Value Function

3.3.1 Adaptive Estimator of Value Function and Θ Matrix Under Two Cases

Case 1: PUs are deactivated

Then, using adaptive control theory [6], the value-function for mth SU in enhanced CRN under Case 1can be represented in vector form as

$$V(E_{m,k}^{SU,1}, \upsilon_{m,k}^{SU,1}) = (z_{m,k}^{SU,1})^T \Theta_{m,k}^{SU,1} z_{m,k}^{SU,1} = (\theta_{m,k}^{SU,1})^T \bar{z}_{m,k}^{SU,1} \tag{8}$$

where $\bar{z}_{m,k}^{SU,1} = vec(z_{m,k}^{SU,1})$, $z_{m,k}^{SU,1} = [(E_{m,k}^{SU,1})^T \upsilon_m^T (E_{m,k}^{SU,1})]^T$ and $\bar{z}_{m,k}^{SU,1} = [(z_{m,k1}^{SU,1})^2, \ldots, z_{m,kn-1}^{SU,1} z_{m,kn}^{SU,1}, (z_{m,kn}^{SU,1})^2]$ is the Kronecker product quadratic polynomial basis vector [7] for mth SU in enhanced CRN under Case 1, $vec(\bullet)$ function is constructed by stacking the columns of matrix into one column vector with off-diagonal elements [7].

Next, the value function of mth SU in Case 1, $V(E^{SU,1}_{m,k}, \upsilon^{SU,1}_{m,k})$, can be approximated by using adaptive estimator in terms of estimated parameter $\hat{\Theta}^{SU,1}_{m,k}$ as

$$\hat{V}(E^{SU,1}_{m,k}, \upsilon^{SU,1}_{m,k}) = (z^{SU,1}_{m,k})^T \hat{\Theta}^{SU,1}_{m,k} z^{SU,1}_{m,k} = (\hat{\theta}^{SU,1}_{m,k})^T \bar{z}^{SU,1}_{m,k} \tag{9}$$

where $\hat{\theta}^{SU,1}_{m,k}$ is the estimated value of mth SU's target parameter vector $\theta^{SU,1}_{m,k}$. It is observed that mth SU's Bellman equation in Case 1 can be rewritten as $J^{SU,1}_{m,k+1}(E) - J^{SU,1}_{m,k}(E) + r(E^{SU,1}_{m,k}, \upsilon^{SU,1}_{m,k}) = 0$. However, this relationship does not hold when we apply the estimated matrix $\hat{\Theta}^{SU,1}_{m,k}$. Hence, using delayed values for convenience; the residual error associate with Eq. (13) can be expressed as

$$\begin{aligned} e_{m,\theta k} &= \hat{J}^{SU,1}_{m,k}(E) - \hat{J}^{SU,1}_{m,k-1}(E) + r(E^{SU,1}_{m,k-1}, \upsilon^{SU,1}_{m,k-1}) \\ &= r(E^{SU,1}_{m,k-1}, \upsilon^{SU,1}_{m,k-1}) + (\hat{\theta}^{SU,1}_{m,k})^T \Delta Z^{SU,1}_{m,k-1} \end{aligned} \tag{10}$$

where $\Delta Z^{SU,1}_{m,k-1} = \bar{z}^{SU,1}_{m,k} - \bar{z}^{SU,1}_{m,k-1}$ is the first difference of regression function of mth SU under Case 1. Next, we define the auxiliary residual error vector as

$$\Xi^{SU,1}_{m,k} = \Gamma^{SU,1}_{m,k-1} + (\hat{\theta}^{SU,1}_{m,k})^T \Delta \mathbf{Z}^{SU,1}_{m,k-1} \tag{11}$$

where $\Gamma^{SU,1}_{m,k-1} = [r(E^{SU,1}_{m,k-1}, \upsilon^{SU,1}_{m,k-1}) .... r(E^{SU,1}_{m,k-1-i}, \upsilon^{SU,1}_{m,k-1-i})]$ and $\Delta \mathbf{Z}^{SU,1}_{m,k-1} = [\Delta Z^{SU,1}_{m,k-1} .... \Delta Z^{SU,1}_{m,k-1-i}]$, $0 < i < k-1$ and $\forall m \in \{SUs\}$. Then the dynamics of auxiliary residual error vector Eq. (11) are generated as $\Xi^{SU,1}_{m,k+1} = \Gamma^{SU,1}_{m,k} + (\hat{\theta}^{SU,1}_{m,k+1})^T \Delta \mathbf{Z}^{SU,1}_{m,k}$.

Now define the update law of the mth SU's time varying matrix $\hat{\Theta}^{SU,1}_{m,k}$ in Case 1 as

$$\hat{\theta}^{SU,1}_{m,k+1} = \Delta \mathbf{Z}^{SU,1}_{m,k} [(\Delta \mathbf{Z}^{SU,1}_{m,k})^T \Delta \mathbf{Z}^{SU,1}_{m,k}]^{-1} [\alpha^{SU,1}_{\theta} (\Xi^{SU,1}_{m,k})^T - (\Gamma^{SU,1}_{m,k})^T] \tag{12}$$

where $0 < \alpha^{SU,1}_{\theta} < 1$.

Next, the estimation of mth SU's optimal power design under Case 1 will be derived based on tuned parameter $\hat{\Theta}^{SU,1}_{m,k}$ as

$$P^{SU,1}_{m,k+1} = \hat{\upsilon}^{SU,1}_{m,k} I^{SU,1}_{m,k} = -(\hat{\Theta}^{\upsilon\upsilon,SU,1}_{m,k})^{-1} \hat{\Theta}^{\upsilon E,SU,1}_{m,k} z^{SU,1}_{m,k} I^{SU,1}_{m,k} \tag{13}$$

Case 2: PUs are activated

Similar to Case 1, the value-function for mth PU and mth SU in CRN under Case 2 can be estimated in vector form respectively as

$$\begin{aligned} l\text{th PU} : \hat{V}(E_{l,k}^{PU,2}, \upsilon_{l,k}^{PU,2}) &= (z_{l,k}^{PU,2})^T \hat{\Theta}_{l,k}^{PU,2} z_{l,k}^{PU,2} = (\hat{\theta}_{l,k}^{PU,2})^T \bar{z}_{l,k}^{PU,2} \\ m\text{th SU} : \hat{V}(E_{m,k}^{SU,2}, \upsilon_{m,k}^{SU,2}) &= (z_{m,k}^{SU,2})^T \hat{\Theta}_{m,k}^{SU,2} z_{m,k}^{SU,2} = (\hat{\theta}_{m,k}^{SU,2})^T \bar{z}_{m,k}^{SU,2} \end{aligned} \tag{14}$$

where $\bar{z}_{l,k}^{PU,2} = vec(z_{l,k}^{PU,2})$, $\bar{z}_{m,k}^{SU,2} = vec(z_{m,k}^{SU,2})$ are Kronecker product quadratic polynomial basis vector for lth PU and mth SU in enhanced CRN under Case 2.

Next, the update law of lth PU's and mth SU's time varying matrices, $\hat{\Theta}_{l,k}^{PU,2}$, $\hat{\Theta}_{m,k}^{SU,2}$, in Case 2 can be derived respectively as

$$\begin{aligned} l\text{th PU} : \hat{\theta}_{l,k+1}^{PU,2} &= \Delta\mathbf{Z}_{l,k}^{PU,2}[(\Delta\mathbf{Z}_{l,k}^{PU,2})^T \Delta\mathbf{Z}_{l,k}^{PU,2}]^{-1}[\alpha_{\theta}^{PU,2}(\Xi_{l,k}^{PU,2})^T - (\Gamma_{l,k}^{PU,2})^T] \\ m\text{th SU} : \hat{\theta}_{m,k+1}^{SU,2} &= \Delta\mathbf{Z}_{m,k}^{SU,2}[(\Delta\mathbf{Z}_{m,k}^{SU,2})^T \Delta\mathbf{Z}_{m,k}^{SU,2}]^{-1}[\alpha_{\theta}^{SU,2}(\Xi_{m,k}^{SU,2})^T - (\Gamma_{m,k}^{SU,2})^T] \end{aligned} \tag{15}$$

where tuning parameter $0 < \alpha_{\theta}^{PU,2} < 1$ and $0 < \alpha_{\theta}^{SU,2} < 1$.

Then, we can derive the adaptive optimal DPA design for lth PU and mth SU in enhanced CRN under Case 2 based on tuned parameter $\hat{\Theta}_{l,k}^{PU,2}$, $\hat{\Theta}_{m,k}^{SU,2}$ respectively as:

$$\begin{aligned} l\text{th PU} : P_{l,k+1}^{PU,2} &= \hat{\upsilon}_{l,k}^{PU,2} I_{l,k}^{PU,2} = -(\hat{\Theta}_{l,k}^{\upsilon\upsilon,PU,2})^{-1} \hat{\Theta}_{l,k}^{\upsilon E,PU,2} z_{l,k}^{PU,2} I_{l,k}^{PU,2} \\ m\text{th SU} : P_{m,k+1}^{SU,2} &= \hat{\upsilon}_{m,k}^{SU,2} I_{m,k}^{SU,2} = -(\hat{\Theta}_{m,k}^{\upsilon\upsilon,SU,2})^{-1} \hat{\Theta}_{m,k}^{\upsilon E,SU,2} z_{m,k}^{SU,2} I_{m,k}^{SU,2} \end{aligned} \tag{16}$$

3.3.2 Closed-Loop Adaptive Optimal DPA System Stability for PUs and SUs in Enhanced CRN

In this section, it will be shown that mth SU's time-varying matrix, $\Theta_{l,k}$, and related value function estimation errors dynamic are asymptotic stable (AS) when PUs in enhanced CRN are deactivated. Further, the estimated adaptive optimal distributed power allocation will approach the optimal power allocation asymptotically. Next the initial system states (i.e. SIR errors of SUs) are considered to reside in the compact set which in turn is stabilized by using the initial stabilizing input $\upsilon_{0m,k}^{SU,1}$ for case 1 and $\upsilon_{m,0}^{PU,2}$, $\upsilon_{m,0}^{SU,2}$ for case 2. Further sufficient condition for the adaptive estimator tuning gain $\alpha_{\theta}^{SU,1}$, $\alpha_{\theta}^{SU,2}$, $\alpha_{\theta}^{PU,2}$ are derived to ensure the all future PUs' and SUs' SIR errors will converge to zero for two cases. Then it can be shown that the actual adaptive DPA approaches the optimal power allocation for PUs and SUs in two cases asymptotically. The algorithm represented the proposed adaptive optimal distributed power allocation is given as follows.

Algorithm 1: Adaptive Optimal Distributed Power Allocation in Enhanced CRN

```
1:  **Initialize:** Adaptive estimator parameters and implementing admissible policy.
2:  **while** $\{ kT_s < t < (k+1)T_s \}$ **do**
3:    **If** Case 1 (i.e PUs are deactivated)
4:        **Calculate** the value function estimation error $\Xi_{m,k}^{SU,1}$.
5:        **Update** the parameters of the value function estimator (12)
6:        **Update** adaptive optimal DPA based on estimated $\Theta_{m,k}^{SU,1}$ matrix (13).
7:    **else** (i.e. Case 2: PUs are activated)
8:        **Calculate** the value function estimation errors $\Xi_{l,k}^{PU,2}, \Xi_{m,k}^{SU,2}$ forPU and SU
9:        **Update** the parameters of the value function estimator (15)
10:       **Update** adaptive optimal DPAbased on estimated $\Theta_{l,k}^{PU,2}$ , $\Theta_{m,k}^{SU,2}$ matrices (16)
11:   **end if**
12: **end while**
13:**Go to** next time interval $[(k+1)T_s, (k+2)T_s)$, and then go back to line 2.
```

4 Numerical Simulations

In this simulation, proposed AODPA is implemented for PUs and SUs in enhanced CRN with channel uncertainties. Since wireless channel attenuations of users in enhanced CRN are different, initial PUs' and SUs' SIRs are different values. Moreover, the augment SIR error system state is generated as $z_k = [E_k \upsilon_k]^T \in$ and the regression function for value function is generated as $\{z_1^2, z_1 z_2, ..., z_2^2, ..., z_3^2\}$ as per Eq. (11). The design parameter for the value function estimation is selected as $\alpha_\theta = 0.0001$ while initial parameters for the adaptive estimator are set to zeros at the beginning.

In Figs. 2 through 3, the performance of proposed adaptive optimal distributed power allocation scheme is evaluated. In Fig. 2, the averages of all PUs' SIRs and SUs' SIRs are shown. Note that proposed AODPA cannot only force SUs converge to low target SIR (i.e. $\gamma_H^{SU} = -10\,\text{dB}$) when PUs are deactivated, but also force PUs and SUs converge to target SIRs (i.e. $\gamma^{PU} = -7\,\text{dB}, \gamma_L^{SU} = -20\,\text{dB}$) respectively while CRN is at Case 2. Also, the power consumptions averages of PUs and SUs are shown in Fig. 3. Obviously, in Case 1, since PUs are deactivated, SUs increase transmission powers to improve network utility (e.g. spectrum efficiency). For Case 2, SUs decrease transmission power to reduce the inference to PUs for guarantying their QoS.

Fig. 2 Average SIRs of PUs and SUs

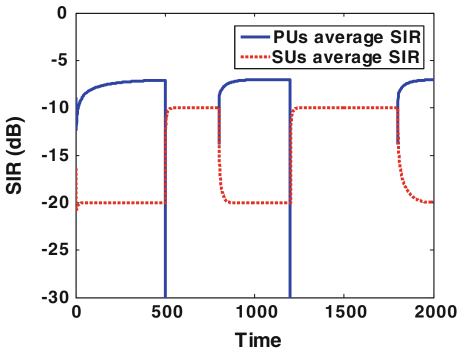
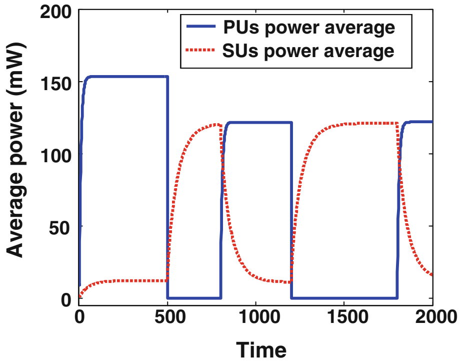


Fig. 3 Average power allocation of PUs and SUs



5 Conclusion

In this work, by using the SIR error dynamics, a novel adaptive dynamic programming scheme is proposed to optimize distributed power allocation (DPA) for both PUs and SUs in enhanced CRN under two cases. The availability of past state values ensured that SIR error dynamic sare not needed for proposed AODPA design while an adaptive estimator (AE) generates an estimated value function and a novel optimal power allocation law based on the estimation of value function. An initial admissible policy ensures that SIR error systems for PUs and SUs in enhanced CRN are stable for two cases while the adaptive estimator learns the value function and the matrix Θ, and optimal power allocation scheme. All adaptive estimator (AE) parameters were tuned online and Lyapunov theory demonstrated the asymptotic stability of the overall closed-loop enhanced CRN system.

References

1. Mitola J, Maguir GQ (1999) Cognitive radio: making software radios more personal. IEEE Pers Commun 6:13–18
2. Haykin S (2005) Cognitive radio: brain-empowered wireless communications. IEEE J Sel Areas Commun 23:201–220
3. Kang X, Liang YC, Arumugam N, Garg H, Zhang R (2009) Optimal power allocation for fading channels in cognitive radio netwokrs: ergodic capacity and outage capacity. IEEE Trans Wirel Commun 8:940–950
4. Wu Y, Tsang HK (2009) Distributed power allocation algorithm for spectrum sharing cognitive radio networks with QoS guarantee. In: Proceedings of 28th IEEE international conference on computer communications, IEEE Press, pp 981–989.
5. Jagannathan S, Zawodniok M, Shang Q (2006) Distributed power control for cellular networks in the presence of channel uncertainties. IEEE Trans Wirel Commun 5:540–549
6. Lewis FL (1999) Optimal control. Wiley, New York
7. Xu H, Jagannathan S, Lewis FL (2012) Stochastic optimal control of unknown linear networked control system in the presence of random delays and packet losses. Automatica 52:1017–1030

Secure Real Time Remote Video Monitoring System

Deepti C. Gavankar and Madhumita Chatterjee

Abstract Video has been an important media for entertainment and communication for many years. Initially video was captured and transmitted in analog form. The digitization of video was done due to the development of computers and digital integrated circuits. The digital video enabled a revolution in the compression and communication of video. With the growth of the multimedia technology video streaming technology is gaining importance. Streaming is simply a technique for transferring data such that it can be processed as a steady and continuous stream. The file is sent to the end user in a (more or less) constant stream. With streaming, the end user can start watching the file almost as soon as it begins downloading. Security becomes a key problem to be handled when valuable multimedia data are travelling over the network. As security of data over the network is a burning issue and dominates the communication systems today, in this paper we present a real time video monitoring security system, named as Secure Video Monitoring System (SVMS) with authentication and selective encryption to securely transfer the valuable multimedia video streams over the un-trusted network. The process of authentication is done by allowing the network access to only authenticated users, data is compressed and decompressed using compression technology. This system provides security by selectively encrypting the streaming data which is send over the network. It also provides video on demand (VOD).

Keywords Video monitoring · Network security · Encryption · Compression

D. C. Gavankar (✉) · M. Chatterjee
Department of Computer Engineering, Pillai's Institute of Information Technology,
Navi Mumbai, India
e-mail: deepti0131@gmail.com

M. Chatterjee (✉)
e-mail: c_a_mita@yahoo.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_37,

1 Introduction

There is a growing need to visually monitor an area from a remote location using a video camera or webcam. The increasing use of streaming media and the proliferation of webcams have improved the ability of users to communicate with each other, to be entertained and to monitor or perform surveillance. Naturally, the development of software for providing such services has become prevalent in the software development community.

The system been developed can be used where security of real time video data while travelling on the network is very important. For example, in banks the manager and his team can keep watch on the employees and the customers by viewing the secured live video. Since the processing done in the banks like counting of cash is sensitive, any unauthorized user should not be able to view it. Even if any unauthorized user hacks the data he will not be able to see the live video since it is encrypted.

2 Problem Definition

A complete video-streaming system involves all of the basic elements of creating, delivering, and ultimately playing the video content. Streaming video is a sequence of "moving images" that are sent in compressed form over the network and are seen by the viewer as they arrive. There exist a diverse range of different video communication and streaming applications which have different operating conditions or properties [1]. Streaming visual data to different users is becoming popular in recent times. One of the main concerns both for the end users and data providers is protecting the transmitted data from every possible security threat. For example, in previous work [2], a video streaming application was developed using H.264 video compression technology, RTP/RTCP as data transfer protocol and Data Encryption Standard (DES) algorithm for security. Since each and every frame in the live video is encrypted, this can result into delay.

The main feature of the suggested design is its ability to provide a secure communication environment for real-time data with minimum delay. So a computer security standard Cipher is used for the selective encryption and decryption of live video data. The live video is sent from the server to the client over the network using TCP/IP Sockets [3]. The data is compressed and decompressed using Motion JPEG (MJPEG) Codec [4]. The software under consideration designs a GUI which allows efficient communication between server and client. The server side maintains a database consisting of various music and video clips that are available for viewing. The server can either stream the secured live video or can display the music or video clip requested by the client. The client side will be able to view either secured live video or requested music or videoclip. Also the client can send a request to the server to upload song or video clip. This request can be a URL or just the name of the song or video clip.

3 Review of Literature

Recent advances in computing and video technology paved the growth for applications and devices in real time video streaming. A survey of technologies for implementing real time video streaming has revealed that it can either be wired or wireless or it can be both. Real Media Real Server, NetShow (next version-Windows Media Technology), NetCamCenter, Webcam Watchdog (next version-Webcam 1-2-3) are some of the real time video streaming software.

3.1 Comparison of Webcam Surveillance Softwares

Vitamin D webcam surveillance software can detect moving objects in surveillance video. It can detect humans in surveillance video and can distinguish them from other objects such as cars. It can send an email when something happens during surveillance video [5]. WebcamXP is webcam surveillance software. It lets connects up to 6 cameras, and view the webcam stream over internet. WebcamXP lets pan/tilt the webcams remotely if the webcams support those. While sharing feed over internet, WebcamXP shows if anyone else is looking at the feed, and shows their IP address, location, and country information. WebcamXP even lets stream the webcam feed to iPhone. Yawcam (Coast Cam is another simple to use home webcam surveillance software. Yawcam stands for Yet Another Webcam Software. As is with other webcam monitoring software, Yawcam comes with Motion Detection and Video Streaming over the internet. Yawcam also supports Stealth mode. In this mode, Yawcam will start whenever your PC starts, and it will immediately start recording without anyone else knowing [6]. NetCamCenter is ideal for monitoring and recording network cameras. It supports various JPEG, MPEG4, H.264 and megapixel IP cameras. Webcam Watchdog also enables to create high quality audio and video for real time streaming.

Real Media RealServer streams audio, video, images, animation, text, and other data types to client computers. RealServer is server software that streams both live and pre-recorded media over a network. The streamed data can originate either on the Internet or within an intranet. The client receives the media in real time, and without having to wait for clips to be downloaded. RealServer uses two protocols for sending instructions and data: Transport Control Protocol (TCP), for sending commands from the client (such as “start” and “pause”) and sending commands from RealServer to clients for specific information (such as the clips’ titles). User Datagram Protocol (UDP) is used for sending actual streamed content [7]. NetShow Server enables Internet Providers and organizations to deliver the highest-quality audio and video across the Internet or enterprise networks. NetShow Services allow users to receive audio and video broadcasts from their personal computers. NetShow Services consist of server and tools components for streaming audio, video over networks. The streaming media components of the Windows Media Technologies, the next version of NetShow Server provide a complete solution for integrating audio and video into online applications, bringing the vibrant power of networked multimedia to the Internet and corporate intranets. The Windows Media Player continuously decompresses and plays the content in real time [7].

All the softwares discussed above focus only on monitoring. Since security of live video data when sent on the network is an important issue, we propose a system in which the live video is selectively encrypted and then sent over the network. Along with this, the facility of Video on Demand (VOD) is also provided.

4 Proposed System

Real time remote video monitoring is becoming increasingly popular in recent times and security of the transmitted data has become one of concerns for the end users and data providers. The objective of this paper is to develop a secured real time video monitoring system, which compresses and selectively encrypts streaming video to enhance the security. The system has an additional feature of video on demand (VOD) where the client can select video clip or music as per their interest. Beside the client can also send a request to the server to upload any music or video clip by giving an URL or just name of song or video clip. The server can directly download the requested music or video clip by using the URL via internet. Then the requested music or video clip can be uploaded fulfilling the client's request.

4.1 System Structure

The system is mainly composed of video capture, compression and selective encryption, client socket network transmission, decryption, decompression and video playback. Considering the real-time video data transmission and security in the software design process, TCP/IP network transmission and data encryption technology Cipher [8] is used.

According to the overall design, logically this system will be divided into two main modules: client and server. The server software must be installed in the host and the client can be installed according to specific needs. A webcam is connected to the server. The systems that are connected to the server over network (LAN) will be the clients which require jdk to be installed in them. A client server type relationship is used where the client request for the service and server fulfills it. The server side is divided into two components the main server components and the database. Multiple clients can be connected to server at any instant. Object oriented programming style is implemented.

Server module is installed in the host and it has the function of being video front-end, is the acquisition and output terminal of the video data; client module can be installed in any host and is the receiving terminal of the video data. The server software and client software are combined into a complete video monitoring system [2].

4.2 Working of the System

System structure (refer Fig. 1) consists of video capturing, compression/decompression using Motion JPEG codec, selective data encryption/decryption using Cipher [9], data transfer over the network using TCP/IP socket and video playback.

The system consists of secure real time video streaming and video on demand (VOD). In real time video, when the client request for live video, the server which has a webcam connected to it starts displaying the live video to the client. In this system the video is not just sent over the network to the client. When the user requests for live video, the server starts capturing the live video using a webcam. The Java Media Framework captures the video data in byte format. This data is compressed using Motion JPEG codec. The compressed data is selectively encrypted using Cipher and send to the client over the network. Selective encryption factor ‘x’ can be used for Cipher. The selective encryption factor ‘x’ is tested for values 3, 5, 7, 10. On the client side the video is decrypted, decoded and displayed to the user.

In VOD the server is intended to act as source of information to be transmitted over the network. Music and video clips are stored in form of files on server. The list of such files is to be presented on client side. The user on client machine is given a choice to select any music or video clip from the list at any moment of time for viewing. Several users can select the same music or video clip at different moments. It must make it possible to load or reload same or different music or video clip across several nodes. The music or video clip selected by user would be displayed on an interactive interface on machine. The user can stop and switch over to other music or video clip on the list any time.

For VOD, the server side has two modules: administrator (SP) and broadcaster (BC). Administrator module creates and control broadcaster module. Number of broadcasters can be created which will handle the client requests. The client can request for video or music which he is interested in. The requested video or music which is available in server database will be then sent to the client. The video or music can also be downloaded and saved on client machine. The client can also send a request which can be name of video or music or it can be a direct URL of that

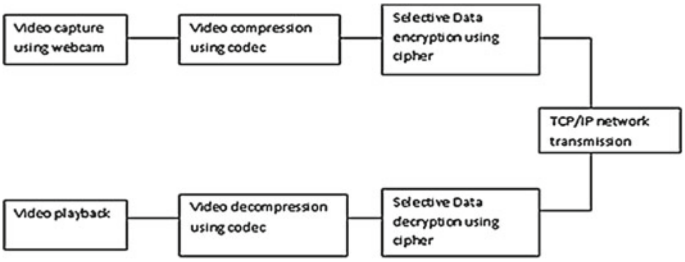

Video capture using webcam
Video compression using codec
Selective Data encryption using cipher
TCP/IP network transmission
Selective Data decryption using cipher
Video decompression using codec
Video playback

Fig. 1 Working of the system

specific video or music. The broadcaster who will receive this client request can upload video or music from the system or from the specific URL.

An effective user friendly interface is required to be designed on client machine for interactive viewing, making the session interesting. Such interface can be designed using jsp. Network connection has to be established properly for easy transmission.

Algorithm for Secure Video Monitoring System.

1. Start up the server.
2. Client opens the web browser and connects to the server by entering the IP address.
3. On authentication the client can request for live video or video on demand (VOD).
4. If the client requests for live video the Java Media Framework on the server captures live video in byte format using a webcam.
5. The video is compressed using Motion JPEG Codec and then it is selectively encrypted using Cipher on the server side.
6. The compressed and selectively encrypted video is sent to the client over LAN by the server.
7. It is decrypted and decompressed and the client receives live video.
8. If the client requests for VOD then the requested music or video is sent to the client over LAN by the server.
9. The client receives requested music or video.

Algorithm for Selective Encryption.

- Java Media Framework (JMF) captures the live video in byte format using webcam.
- The frames are compressed using MJPEG codec.
- Every ‘x’ bit is encrypted using a key.
- ‘x’ is bit number which can be changed as per requirement. For e.g. encrypting every 10th bit of image

Algorithm for Compression.

- Java Media Framework (JMF) captures the live video in byte format using webcam.
- Motion JPEG initially divides each frame into number of macro blocks. Each macro block contains 4 Y (luminance) block, 1 Cb (blue color difference) block, 1 Cr (red color difference) block (4:2:0) in the frame which are used for detecting motion, encoding and producing output motion JPEG frame [10].
- Macro block contains RGB data which is converted into JPEG.
- In order to discover motion the current frame is compared with previous frame.

5 Implementation

Implementation of any software is always preceded by important decisions regarding selection of the platform, the language used etc. These decisions are often influenced by several factors such as real environment in which the system works the security concerns and other implementation specific details. In the proposed system, the operating system used is Microsoft Windows XP. The Windows XP provides a suitable environment for the smooth functioning of the project. Tomcat 5.5 is used as application server [11] and Ms-Access as database. The developing tool is Java (jdk1.6.0_22). Java media framework (JMF) is used for handling streaming media in java programs [12]. Hyper Text Transport Protocol (HTTP) is used for VOD. Internet Explorer is used as web browser.

6 Result

The proposed system provides secured way to transfer live video from server to client over the network. The security is provided by compressing and selectively encrypting the live video while it is been sent over the network. It also provides the facility of video on demand (VOD). This section shows result for the proposed Secured Video Monitoring System (SVMS).

Figure 2 shows GUI for SVMS where the client requests for live video data from the server. When the user requests the live video the server starts capturing the live video using a webcam, encodes and encrypts it and sends it to the client. On the client side the video is decrypted, decoded and displayed to the user.

Figure 3 shows GUI for SVMS where the live video is not seen on the screen. This happens when an unauthorized user hacks the data over the network and tries to view it. Since the user is not authorized he does not have the decryption code to decrypt the encrypted data which runs on the client side. So he just sees a blank window instead of the live video.

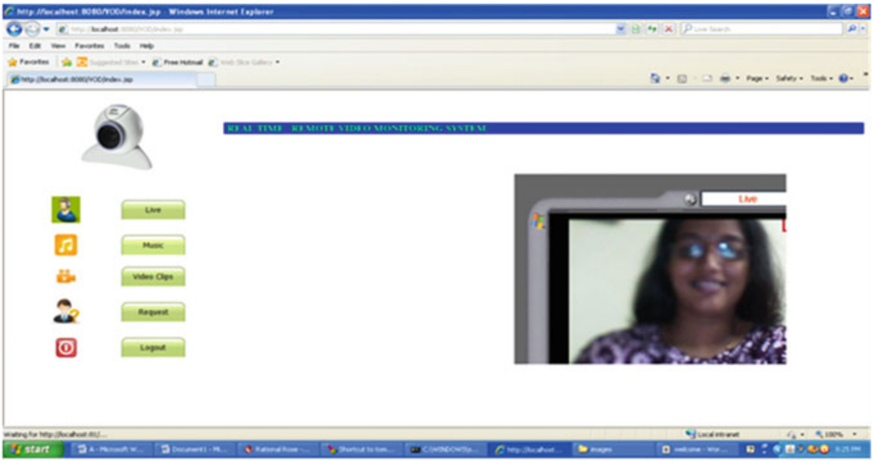

http://localhost:8080/VOD/index.jsp - Windows Internet Explorer
REAL TIME REMOTE VIDEO MONITORING SYSTEM
Live
Music
Video Clips
Request
Logout
Live
start

Fig. 2 Secured live video

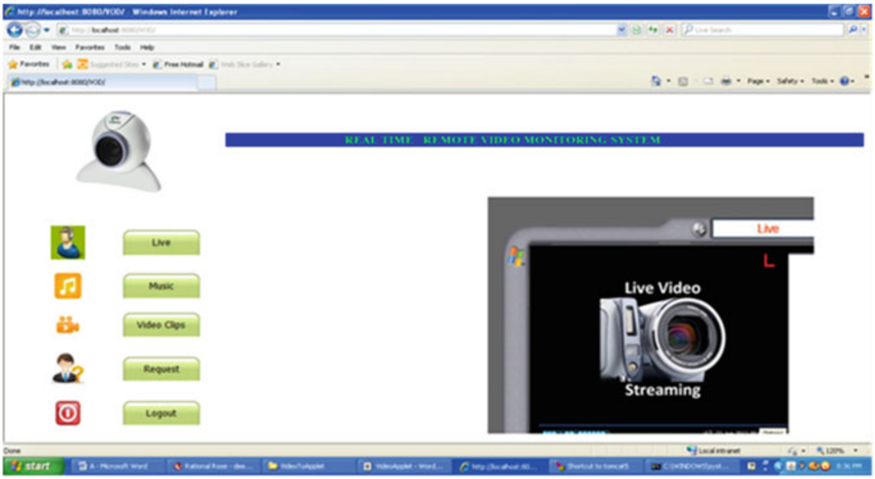

http://localhost:8080/VOD/ - Windows Internet Explorer
File Edit View Favorites Tools Help
Favorites
Free Hotmail
http://localhost:8080/VOD/
Page Safety Tools
REAL TIME REMOTE VIDEO MONITORING SYSTEM
Live
Music
Video Clips
Request
Logout
Live
L
Live Video
Streaming
Done
Local intranet
120%
start

Fig. 3 Non-streaming of live video

7 Analysis

Secure Video Monitoring System (SVMS) that we have proposed is a secured real time video monitoring system which is composed of server host, client host and webcam. As network data security is an important issue and dominates the communication systems today, we present a security scheme, named as SVMS (Secure Video Monitoring System) to securely transfer the valuable multimedia video streams on the un-trusted network using the authentication and selective encryption scheme.

The process of authentication is done by allowing the network access to only authenticated users. Here the data is compressed and decompressed using compression technology called MJPEG Codec. Cipher is used for selective encryption and decryption so that the video can securely travel over the network. So even if any unauthenticated user hacks the data he will not be able to see the live video since it is encrypted. Since selective encryption is used, there will be minimum delay and the user can see the real time live video. In previous work [2], the system provides encryption using Data Encryption Standard (DES) on every frame. This may introduce delay which minimizes the meaning of real time.

The system also provides video on demand (VOD) where the user can request for music and video clips. The user can also send a request to the server to upload new music or video clips. This request can be sent in form of URL or just as name of the music or video clip. If the request is sent in form of URL, the server can directly download the requested music or video clip from the URL using internet.

Table 1 shows comparative analysis of Real Time Remote Video Surveillance Systems where different softwares are compared on parameters like compression technique, protocol, number of cameras, encryption technique and VOD facility.

Table 1 Comparative analysis of real time video surveillance systems

Softwares\ Features	Compression technique	Protocol	No. of cameras	Encryption technique	VOD facility
Real Media Real Server	NA	TCP/UDP	01	No	No
NetShow	MPEG-4	RTSP	01	No	No
NetCamCenter	H.264	RTSP	01	No	No
Webcam Watchdog	MJPEG	TCP/IP	01	No	No
Vitamin D	MPEG-4	RTSP	01	No	No
WebcamXP	MJPEG	RTSP	06	No	No
Yaw cam (Coast cam)	MJPEG	TCP/IP	01	No	No
Secure Video Monitoring System	MJPEG	TCP/IP Sockets and HTTP	01	Cipher	Yes

8 Conclusion

Video monitoring witnessed lot of work in past few years, but it is observed that least emphasis was given on security of data travelling through the network. In this paper we have developed an environment with security infrastructure that performs secure video monitoring. The compression and selective encryption technique are used in such a way that they provide a security solution to the data travelling upon the network. It is observed that even after much data processing the video is displayed with minimum delay. It also provides good quality of video even after data processing. A security based real-time remote video monitoring system along with video on demand (VOD) is implemented, with the specific implementation of the server and client terminal.

The proposed system is a step towards provable security of real world implementations and will be a motivating point for further research in the field of multimedia security. The flexibility of this application further allows the implementation of other key management and encryption algorithm.

References

1. Apostolopoulos JG et al. (2002) Introduction. In: Video streaming: concepts, algorithms, and systems. Mobile and Media Systems Laboratory, HP Laboratories, Palo Alto. HPL-2002-260, 18th Sept 2002
2. Liu Y et al. (2010) Research on real-time remote video monitoring system. In: The 2nd international conference on computer and automation engineering, Singapore, pp 484–487
3. Calvert KL, Donahoo MJ (2008) Basic sockets. In: TCP/IP sockets in Java: practical guide for programmer, 2nd edn. MKP, Burlington, pp 15–26
4. Harte L (2012) IPTV basics. http://www.scribd.com/doc/48288310/94/Motion-JPEG-MJPEG
5. Rob H (2010) Vitamin D. http://www.vitamindinc.com/features_smart.html
6. M Lundvall (2012) Yawcam. http://www.yawcam.com
7. CSWL labs basic streaming technology and RTSP protocol. http://www.cswl.com

8. Ibrahim AAK (1991) Cryptography and data security: cryptographic properties of arabic. In: Proceedings of the third Saudi engineering conference, Riyadh, Saudi Arabia, vol 2, pp 910–921, 24–27 Nov 1991
9. Churchhouse RF (2002) Introduction. In: Codes and ciphers Julius Caeser, the Enigma and the internet, 1st edn. CUP, Cambridge, pp 5–6
10. Bohoris C (1998) Simulation and optimization of JPEG and motion JPEG. MSc Telematics (Telecommunications and Computer Engineering), Project report, University of Surrey, UK.
11. Turk M (1999) The Apache Software Foundation. http://www.apache.org/
12. Singh G (2006) Secure video conferencing for web based security surveillance system. MS Thesis, Dept. Comp Sci. and Eng., IIT, Kanpur, July 2006

Fully Self-organized Key Management Scheme in MANET and Its Applications

Fuyou Miao, Wenjing Ruan, Xianchang Du and Suwan Wang

Abstract The paper first proposes a fully self-organized key management scheme in mobile ad hoc networks which is both certificateless and free from any trusted third party such as Certificate Authority or Key Generation Center. The scheme allows a node to set up the public/private key pair all by itself and use the public key as its identity according to the property of ad hoc networks. Based on the scheme, some applications such as Encryption, Signature, Signcryption algorithms are given. Among these algorithms, the security of encryption algorithms is detailed and the AdvancedEnc is proved to be IND-CCA secure encryption algorithm under the random oracle model. These applications show that our key management scheme is self-organized, simple, efficient and practical.

Keywords Mobile network · Ad hoc · Key management · Fully self-organized · Random oracle model

1 Introduction

Mobile ad hoc networks (MANETs) are wireless mobile multi-hops networks which do not rely on any fixed infrastructure, and can be widely used in the battlefield, counter-terrorism, emergency rescue, as well as a variety of occasions which lack of fixed network. It is a necessary complementation and extension to the Internet and it has also been one of the hotspots of network research in recent years.

F. Miao · W. Ruan (✉) · X. Du
School of Computer Science and Technology, University of Science
and Technology of China, He Fei, China
e-mail: ruanwj@mail.ustc.edu.cn

S. Wang
School of Computer Science and Technology, University of Anhui, He Fei, China

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_38,

In addition to mobility, self-organized, MANETs also tend to have such characteristics as full distribution, and temporary. Specifically, there aren't any fixed infrastructures to provide services for nodes in the networks. All nodes are equal and often need to collaborate to complete a task; each node is completely independent to manage and control their own resources and determine their own behaviors; MANETs tend to set up for a particular purpose or temporary emergency. Mostly due to that the nodes in the network are mobile devices and their energy is often limited, so these nodes may not exist for a long time.

Similar to Internet, security in MANET is a critical problem. For instance, messages sent from a source to the destination often need to be confidential; digital signature is often required to ensure the integrity and right source of a message. All these security mechanisms are based on key management in MANETs, which provides key generation and distribution for them.

However, the traditional certificate-based key management scheme PKI needs a fixed Certificate Authority to manage certificates for all nodes; moreover, the certificate-based scheme uses a certificate to bind the public key and identity of a node which results in the complication and inefficiency of key management. The identity based key management scheme, although use identity as the public key and is of no certificate, needs a Private Key Generator as the trusted third party. Therefore, neither PKI nor identity based scheme fits for the mobile ad hoc networks, which is characterized by self-organization, distribution and autonomy.

Thus, allowing for the above properties of MANET, the paper proposes a fully self-organized key management scheme, which is both certificateless and free from any trusted third party such as Certificate Authority or Key Generator Center. The scheme allows a node to set up the public/private key pair all by itself and use the public key as its identity according to the property of ad hoc networks. Besides, some applications and proofs are given to show the practical availability.

The paper is organized as follows: In part 2, related work gives the current related mechanisms, their advantages and disadvantages; it is followed by the preliminary knowledge in part 3. Part 4 describes the specific our scheme and its applications in detail. Part 5 deals with the security proof of encryption application. At last, a brief summary of our work and some future work is given in part 6.

2 Related Works

For certificate-based key management mechanism [1], there is a Certificate Authority (CA) which is responsible for certificate issue, update and revocation. However, different from the traditional fixed networks, mobile ad hoc networks are of the above mentioned characteristics: distribution, self-organization, autonomy. So, it is difficult to designate a node as such a CA.

Authors of papers [2–4] introduce a (t, n) threshold public key cryptography in MANET in which the CA is composed of n servers. Many public key certificate management operations are completed by at least t out of n servers together. Obviously, its efficiency becomes low and some of these server nodes are easily exploited by adversaries. Many other similar schemes have the same problems.

Miao Fuyou [5] proposed a one-way hash chain based on PKI for MANETs without an online trusted third party. Through a node can initialize, refresh, and revoke its certificate just by itself, but the length of the one-way hash chain is predetermined; the largest number of update cycles is limited. Over a period of time, the one-way hash chain must be changed.

In 1984 Shamir proposed the concept of ID-based cryptography [6, 7] in which the public key can be arbitrary string. It has more advantages compared with certificate-based key management schemes. User's public key can be a hash value of user's name, address, E-mail, identity card number. The user's private key is generated by the Key Generation Center (KGC). Identity-based cryptography allows any two users can securely communicate, and need not to exchange public key certificates or save a list of public key certificates.

The above mechanism doesn't fit for MANET. Firstly, for fully self-organized mobile network model, the existence of KGC is impractical. For individual nodes, it is difficult to establish or vote a fully trusted third party through a good mechanism. Secondly, before a node joins the network it must communicate with the KGC, and its private key generation must rely on the KGC. KGC maintenance will take a lot of manpower and material resources. Such mechanism is of the key escrow problem. KGC knows all nodes' private keys, so if it is captured by an attacker or it takes the initiative to make malicious behaviors, the consequences will be devastating. Thirdly, the node private key during transmission process is easy to be intercepted, and if so, the node intercepting other's private key can impersonate other nodes in the network to communicate. As a result, some researchers have proposed the concept of distributed KGC (D-KGC) [8, 9]. In the initial stage, choose n nodes as master nodes and generate a secret share for each master node by the mechanism of (n, t) threshold. Thus any at least t master nodes can recover the master key. Each new node's private key generation must be completed by at least t master nodes rather than a single KGC. This solves the problem of key escrow, but it makes network traffic largely increase to communicate with at least t master nodes. What's more, the secure transmission of private key shares is still a problem.

3 Preliminaries

3.1 Notation

To better understand the main idea of our system, we first list the main notations used in the paper in Table 1.

Table 1 Main notations used in our system

Notations	Description
p, q	Two large primes
G_1	A q-order subgroup of the additive group of points of E/Fp
G_2	A q-order subgroup of the multiplicative group of the F^*_{p2}
e	Pairing s.t. $G_1 \times G_1 \rightarrow G_2$
H_1	A hash function: $G_2 \rightarrow \{0, 1\}n$
H	A hash function: $\{0, 1\}^* \times G_2 \rightarrow Zq^*$
P	Generator of G_1
Q	$\in G_1$, another point of E/Fp
ID_A	Network ID of node A
x_A	Secret key of node A
PA_{pub}	Public key of node A

3.2 Consideration on Notation of Nodes in MANET

It is well known that nodes in MANET are equal, and each node is both a host and a router. Unlike the identity of IP address in Internet which consists of network address and host address, the identity of a MANET node is often unstructured and used just to distinguish one node from the others. Then we can designate the public key as the identity of a MANET node, If we must distinguish between different networks, other parts can be added to the public key as the whole identity. In fact, what we need is to let the identity include the public key of a node.

Although the identity appears to be complex, we must recognize the fact that: the identity of a user is stored in the computer system and the communications among users are largely dependent on the computer system. So a public key can be used as the user's identity.

We consider a fully self-organized mobile ad hoc network without a trusted third party as the system administrator or network planner. A node joins and leaves the MANET dynamically and autonomously. There are only one or several nodes in the initial network. Subsequent nodes can freely join and leave the network. A new node which wants to join this network can select its secret key and compute the public key all by itself. Next it deals with the public key as its own identity. Since a public key is used as the identity of a node, getting the identity of a node means obtaining the public key.

3.3 Mathematical Problems

Bilinear Pairing: suppose G_1 and G_2 are an additive cyclic group and multiplication cyclic group of the same prime order q respectively, so that let bilinear pairings e: $G_1 \times G_1 \rightarrow G_2$ satisfy the following properties:

1. Bilinear: $e(aP, bQ) = e(P, Q)^{ab}$ for all $P, Q \in G_1$ and $a, b \in Zq^*$.
2. Non-degenerate: exist $P\in G_1$and $Q\in G_1$, make $e(P, Q) \neq 1$.
3. Computability: There is an effective algorithm to compute e(P, Q), for any P, $Q \in G_1$.

The Bilinear Diffie-Hellman Problem (BDH): Let G_1, G_2 be an additive cyclic group and multiplication cyclic group of prime order q and e: $G_1 \times G_1 \rightarrow G_2$, be an admissible bilinear map. Given {P, aP, bP, cP}, to compute $e(P, P)^{abc}$ is called Bilinear Diffie-Hellman Problem (BDH), where P is a generator of G_1, a, b, c are random values in Zq*. The BDH problem is supposed to be intractable at present and is one of the foundations for constructing a public key scheme.

4 Fully Self-Organized Key Management Scheme

4.1 Public Key Scheme

As mentioned above, the identity of an ad hoc network node could be the IP/MAC address, email address or any other information that can be used to identify the node uniquely. Thus we can designate the public key as the identity of an ad hoc node.

Setup: Assume all nodes in an ad hoc network share the parameters {Q, P}, where $P, Q \in G_1$, G_1 is a cyclic additive group with the prime order of q; P is a generator of G_1, G_2 is a cyclic multiplicative group with the same order q, e is a bilinear pairing: $G_1 \times G_1 \rightarrow G_2$;

Key Generation: Before joining an ad hoc, a node A first selects by itself a private key $x_A \in Zq^*$, computes public key $PA_{pub} = x_AP$, and specifies PA_{pub} as its identity. Thus the key pair of node A is (x_A, PA_{pub}). Obviously, (x_A, PA_{pub}) is different from an identity based key pair, a node once obtains PA_{pub} , the identity of A, it knows the public key of A as well.

Consideration on Key Update/ Revocation: As mentioned above, a mobile ad hoc network is often temporary and nodes exist just for a short period of time; moreover, the public key is used as the identity of a node; therefore, the public key of a node should keep unchanged once it joins in the network. In fact, if a node updates its public key (identity) silently or leaves the network [10, 11], it means the node revokes its old public key (identity).

The proposed public key scheme has many advantages. (1) Efficient and Simple: Compared to the traditional certificate-based key management schemes, the proposed scheme is free from all the certificate operations such as distribution, validation, update and revocation; in contrast with ID-based schemes, the proposed scheme needs not to communicate with the KGC. That is, the proposed scheme is a light-weight one. (2) Self-organized: on the one hand, the traditional certificate-based key management relies on the support of CA online. On the other hand, although ID-based key management scheme can directly acquire the public key of a node according to its identity information, but the calculation of the private key must rely on KGC. Once the KGC is comprised, the security of the whole system will collapse. But the proposed scheme does not need any support of TTP and all key management operations are done by the node itself.

4.2 Some Applications

Next, we will give some applications of the proposed key management scheme to show the availability. Among these applications we detail the encryption algorithm AdvancedEnc and prove it to be IND-CCA secure.

4.2.1 BasicEnc: An IND-CPA Secure Encryption Algorithm

Encrypt: if node A wants to send a message m to node B (which holds the private key $\mathrm{x_B}$, and the public key $\mathrm{PB}_{pub} = \mathrm{x_B P}$), it does as follows:
A first selects r∈ Zq* at random, computes:

$$\begin{aligned} \mathrm{Q_{prt}} &= \mathrm{rQ} \\ \mathrm{U} &= \mathrm{rP}, \text{ keeps r and } \mathrm{Q_{prt}} \text{ in private,} \\ \mathrm{V} &= \mathbf{m} \oplus \mathrm{H_1(e(PB}_{pub}, \mathrm{Q_{prt}))} \end{aligned} \tag{1}$$

Thus {U, V} is the ciphertext C that A sends to B.
Decrypt: Receiving the ciphertext C, B decrypts it as follows to get the original message:

$$\begin{aligned} \mathbf{m} &= \mathrm{V} \oplus \mathrm{H_1(e(U, x_B Q))} \qquad (2) \\ &\quad \text{for} \mathrm{V} \oplus \mathrm{H_1(e(U, x_B Q))} \\ &= \mathrm{V} \oplus \mathrm{H_1(e(rP, x_B Q))} \\ &= \mathrm{V} \oplus \mathrm{H_1(e(x_B P, rQ))} \\ &= \mathrm{V} \oplus \mathrm{H_1(e(PB}_{pub}, \mathrm{Q_{prt}))} \\ &= \mathbf{m}; \end{aligned}$$

We can simply find from the above equation that only node A and node B have the ability to resume the message m from C.

4.2.2 AdvancedEnc: An IND-CCA Secure Encryption Algorithm

Now in order to improve the security of BasicEnc to make it to be an IND-CCA2 secure (IND-CCA for short), we modify it as follows according to paper [12].

Let SynEnc be a symmetric encryption algorithm ($C = a \oplus M$; $M = C \oplus a$). Let the modified encryption algorithm be called Epk' described as follows:
$\text{AdvancedEnc}(M) = \{E_{pk}(\theta, H_2(\theta, M)); M \oplus H_3(\theta)\}$

Let n be the size of plaintext M, H_2 and H_3 are two hash functions, $H_2 : \{0, 1\}^n \times \{0, 1\}^n \rightarrow Zq^*$, $H_3 : \{0, 1\}^n \rightarrow \{0, 1\}^n$.

Key extraction: same as before.

Encrypt: select θ in $\{0, 1\}^n$ randomly, computes $r = H_2(\theta, M)$, the cipher text is

$$\begin{aligned} C &= \{rP, \theta \oplus H_1[e(PB_{pub}, rQ)], M \oplus H_3(\theta)\} \\ &= \{U, V, W\} \end{aligned}$$

Decrypt: ① $\theta = V \oplus H_1[e(x_B Q, U)]$
② $M = W \oplus H_3(\theta)$
③ if $r = H_2(\theta, M)$, accepted; else rejected.

4.2.3 Signature Protocol

Sign: if node A needs to sign a message **m**, it can do as follows:
A selects a random number $k \in Zq^*$ and computes $P_{prt} = rP$ and $Q_{pub} = rQ$, to sign m, it does as follows:

$$\begin{aligned} r &= e(x_A Q, P)^k \\ v &= H(m, r) \\ U &= (v + k)x_A Q \end{aligned} \quad (3)$$

Thus the signature of m is {U, v}.

Verify: If the verifier, e.g. node B, wants to verify the signature {U, v}, it computes:

$$r = e(U, P)e(vQ, -PA_{pub}) \quad (4)$$

And checks if $v = H(m, r)$ holds, if it does, the verification succeeds, or else it fails. That is because:

$$\begin{aligned} r &= e(U, P)e(vQ, -PA_{pub}) \\ &= e((v + k)xAQ, P)e(vQ, -xAP) \\ &= e(k\, xAQ, P) \\ &= e(xAQ, P)k \end{aligned}$$

The above procedure shows that only A has the ability to produce the signature because the private xA is used during the signing.

4.2.4 Signcryption Protocol

Sign and encrypt: if A needs to send a message m to B with confidentiality and integrity guaranteed, it computes c as the Signcryption:

$$\mathrm{c} = \mathrm{m} \oplus \mathrm{H}_1(\mathrm{e}(\mathrm{x}_\mathrm{A}\mathrm{Q}, \mathrm{PB}_{pub}))$$

Verify and decrypt: After receiving c from A, node B resumes m as follows:

$$\mathrm{m} = \mathrm{c} \oplus \mathrm{H}_1(\mathrm{e}(\mathrm{x}_\mathrm{B}\mathrm{Q}, \mathrm{PA}_{pub}))$$

5 Proof of Security for Our Encryption Mechanism in Random Oracle Model

In order to show the availability of the proposed fully self-organized key management scheme, we first prove the BasicEnc is IND-CPA secure and then prove that the AdvancedEnc is IND-CCA secure [13–15].

Lemma 5.1 Let H_1 be a random oracle: $\mathrm{G}_2 \to \{0, 1\}^\mathrm{n}$. n is the space size of plaintext and k is security parameter. Suppose A is an IND-CPA adversary with advantages ε (k) against BasicEnc and makes $\mathrm{Q}_{\mathrm{H}1}$ queries to H_1 in total. Then there exists an algorithm B which can solve the BDH problem for G with advantage at least $2\,\varepsilon(\mathrm{k})/\mathrm{Q}_{\mathrm{H}1}$ and a running time O (time (A)).

Proof Assume CH is the BDH challenger and it always responds to B. The BDH parameters $\{\mathrm{q}, \mathrm{G}_1, \mathrm{G}_2, \mathrm{e}\}$ is produced by G and a random instance of BDH problem is {P, aP, bP, cP}, where a, b, c are randomly selected in Zq* and P is random value in G_1. Now let the BDH solution is $\mathrm{D} = \mathrm{e}(\mathrm{P}, \mathrm{P})^{\mathrm{abc}}$. CH sends {P, aP, bP, cP} to B, next B will find out the D through A.

Setup: Algorithm B creates $\mathrm{KEY}_{pub} = \{\mathrm{q}, \mathrm{G}_1, \mathrm{G}_2, \mathrm{e}, \mathrm{n}, \mathrm{PK}_{pub}, \mathrm{Q}, \mathrm{H}_1\}$, where $\mathrm{PK}_{pub} = \mathrm{aP}, \mathrm{Q} = \mathrm{bP}$, and H_1 is a random oracle controlled by B. then B can be a challenger of A and sends the KEY_{pub} to adversary A. In fact we can see that the corresponding private key is a. But only CH knows it.
Query to $\mathbf{H}_1$: the adversary A can issue a query to the random oracle H_1 at any time. Algorithm B will do the followings to respond to a query Q_i:

1. If Q_i has already appeared before, B will return the same result as the last denoted by $\{\mathrm{Q}_\mathrm{i}, \mathrm{H}_\mathrm{i}\}$;
2. Else then B selects a random number $\mathrm{H}_\mathrm{i} \in \{0, 1\}^\mathrm{n}$ to respond to A and records $\{\mathrm{Q}_\mathrm{i}, \mathrm{H}_\mathrm{i}\}$ in the local.

Adversary A prepares two messages $\{\mathrm{M}_0, \mathrm{M}_1\}$ to B which are to be challenged and B sends them to CH. Then CH selects {P, aP, bP, cP} and send it to B. B returns

$C = M_b \oplus r$ to A, r is selected randomly; $b = 0$ or 1. And B sends ciphertext $C' = \{cP, C\}$ to A. by our definition, $Mb = R \oplus H_1[e(cP, aQ)] = R \oplus H_1$ [D].
Guess: Now A guesses $b' = 0$ or 1. According to the assumption, the correct probability for $b' = b$ is $\varepsilon + 1/2$. Then B selects a $\{\mathbf{Q}_i, H_i\}$ randomly from the local records and sends it to CH as the solution of BDH problem.

Since B provides a real attack environment for A, the probability that A comes up with a query for $H_1(D)$ among Q_{H1} queries (such an event is denoted by E) in above process is the same as in the real environment. Next, we will prove $\Pr[E] \geq 2\,\varepsilon$.

In the real attack, if A never issues a related query and just guesses b′ by intuition, it will success with the advantage $\Pr[b' = b|\neg E] = 1/2$.
Because

$$\Pr[b' = b] = \Pr[b' = b|\neg E] \cdot \Pr[\neg E] + \Pr[b' = b|E] \cdot \Pr[E]$$

And

$$0 \leq \Pr[b' = b|E] \leq 1$$

$$\Pr[b' = b] \geq 1/2 + \varepsilon$$

So

$$1/2 + \varepsilon \leq \Pr[b' = b] \leq \Pr[b' = b|\neg E] \cdot \Pr[\neg E] + \Pr[E]$$

Namely

$$1/2 + \varepsilon \leq \Pr[b' = b] \leq (1/2) \cdot \Pr[\neg E] + \Pr[E]$$
$$1/2 + \varepsilon \leq \Pr[b' = b] \leq (1/2)(1 + \Pr[E])$$

Then it follows

$$\Pr[E] \geq 2\,\varepsilon$$

Now we can know that among the total Q_{H1} queries, such an event E appears with the advantage at least 2ε.Thus, B will select $\mathbf{Q_i}$ from the local records as the response to CH with the correct probability at least $2\varepsilon / Q_{H1}$.

According to the above proof process, it follows that if there is an IND-CPA adversary A against BasicEnc, also we can find out an algorithm B which can solve the BDH problem with the non-negligible probability of at least $2\varepsilon / Q_{H1}$.

So the proposed fully self-organized key management scheme can be used to construct an IND-CPA secure encryption algorithm.

Theorem 5.2 (Chosen-Ciphertext Security) suppose that BasicEnc is a γ-uniform (γ is a probability value) asymmetric one-way encryption algorithm with (ε_1, t_1) and n is the size of plaintext M. Then the AdvancedEnc is IND-CCA secure in the random oracle model with (ε, t). Suppose the adversary of AdvancedEnc issues Q_d decryption queries, Q_{H2} queries to H_2 and Q_{H3} queries to H_3. (See Theorem 14 in paper [12])

$$\varepsilon = (2(Q_{H2} + Q_{H3})\varepsilon_1 + \varepsilon_2 + 1)(1 - 2\varepsilon_1 - 2\varepsilon_2 - \gamma - 2^{-n})^{-Qd} - 1$$
$$t = \min(t_1, t_2) - O((Q_{H2} + Q_{H3}) \cdot n)$$

For the modified encryption algorithm AdvancedEnc, θ, H_2 and H_3 are additional. We have already proved that BasicEnc is IND-CPA secure. To prove the Theorem 5.2, we come up with the Theorem 5.3.

Lemma 5.3 suppose that BasicEnc is a γ-uniform (γ is a probability value) asymmetric IND-CPA encryption algorithm in random oracle. Then it must be a γ-uniform (γ is a probability value) asymmetric one-way encryption algorithm under chosen plaintext attack.

Proof One-way encryption is weaker security notion compared to IND-CPA encryption. In the sense of OWE, let A is an adversary with public key *pk* and cipher text *C*. A can never get decryption services from anywhere and can just select a random message *m* from plaintext space to break the ciphertext with the successful probability of $1/2^n$ if n is the bit length of the plaintext.

However, beside the same prerequisites with OWE, an IND-CPA adversary B can also interact with its challenger and access the random oracles. That is to say, the IND-CPA adversary B is more capable than the OWE adversary in breaking ciphertexts. In order to distinguish between the M_0 and M_1, A has two means: (1) randomly guesses Mb from the two plaintexts; (2) Directly cracks the ciphertext to find out the plaintext. The successful probability of (1) is 1/2, and that of (2) relies on queries to the random oracle. Therefore, ignoring (1), we think that the successful probability for B in (2) must be greater than A in the sense of OWE since they have the same target to break a ciphertext. So if the BasicEnc is IND-CPA secure, it is certainly a OWE algorithm.

Thus the Theorem 5.2 can be proved to be correct by Lemma 5.3, it in turn follows that our modified encryption algorithm AdvancedEnc is IND-CCA secure in the random oracle model with a non-negligible advantage.

6 Conclusions

Since the attributes of full organization, distribution and temporary existence in mobile ad hoc networks, Traditional certificate based key management schemes and ID-based scheme can not fit for the network. Therefore, we propose a fully self-organized key management scheme, which is efficient, simple and TTP-free. To demonstrate the availability, some algorithms including encryption, signature, signcryption are given. Among these algorithms, the paper focuses on the security of encryption algorithms and proves the AdvancedEnc to be IND-CCA secure, which shows that our fully self-organized key management scheme is practical. In fact, we can also construct the practical signature, signcryption and authentication algorithms and prove them to be practical in security.

Acknowledgments The paper is supported by NSF of China (granted No. 60970128, No. 61170233, No. 61272472), Youth Innovation Foundation of USTC of 2010 and China Postdoctoral Science Foundation (No. 2011M501397).

References

1. Hegland AM, Winjum E (2006) Survey of key management in ad hoc networks. IEEE Commun Surv Tutor 8(5):48–66
2. Munivel E, Ajit GM (2010) Efficient public key infrastructure implementation wireless sensor networks. ICWCSC. doi:10.1109
3. Nisbet A, Rashid MA, Alam F (2010) The quest for optimum server location selection in mobile ad hoc networks utilising threshold cryptography. In: International conference on information technology: new generations, pp 891–896
4. Yan X, Fuyou M (2003) Secure distributed authentication based on multi-hop signing with encrypted signature functions in mobile ad hoc networks. Acta Electron Sinica 31(2):161–162
5. Miao Fuyou, Xiong Yan (2005) PKIM: a public key infrastructure for mobile ad hoc network. Chin J Electron 14(4):594–598
6. Shamir A (1985) Identity-based cryptosystems and signature schemes. Adv Cryptol 196:47–53
7. Li L, Wang Z, Liu W, Wang Y (2011) A certificateless key management scheme in mobile ad hoc networks. In: International conference on, wireless communications, pp 1–4
8. Li L-C, Liu R-S (2010) Securing cluster-based ad hoc networks with distributed authorities. IEEE Trans Wirel Commun 9(10):3072–3081
9. Lee H, Kim J, Son J, Kim S, Oh H (2012) ID-based key management scheme using threshold decryption for OPMD environment. IEEE ICCE-2012, pp 733–734
10. Wang Y, Zou X (2007) An efficient key revocation scheme for wireless sensor networks. IEEE ICC, pp 1260–1265
11. Li L, Wang Z, Liu W, Wang Y (2011) A certificateless key management scheme in mobile ad hoc networks. Int Conf Wirel Commun, pp 1–4
12. Fujisaki E, Okamoto T (1999) Secure integration of asymmetric and symmetric encryption schemes. In: Advances in cryptology-crypto '99, Lecture notes in computer science, vol 1666. Springer-Verlag, pp 537–554
13. Boneh D, Franklin M (2001) Identity-based encryption from the weil pairing. Comput Sci 2139:213–229
14. Katz J, Lindell Y (2008) Introduction to modern cryptogarphy: principles and protocols. National Defense Industry Press, China
15. Bellare M, Desai A, Pointcheval D, Rogaway P (1998) Relations among notions of security for public-key encryption schemes. Comput Sci 1462:26–45

A Hierarchical Color Net Model for Smart Grid Security Monitoring

Debraj Ghosh and Nabendu Chaki

Abstract Smart grids equipped with highly valuable components like intelligent electronic devices (IED), smart meters, etc. are replacing the traditional power grids. Smart grids are to depend heavily on ad-hoc, wireless networks and internet for communication. Hence these are quite prone to various types of physical and cyber intrusions. Meter tampering is also treated as intrusion. Intrusion on smart meters has become easier in the sense that no physical tampering of the device is required. Instead a smart hacker can corrupt the data being gathered and transmitted electronically. In this paper, a hierarchical Color net model has been proposed for power transmission and distribution from the main power station to the consumers. The proposed model detects any tampering of the smart meter data that occurs at consumer level and carries the security information (affected smart meter ids) to the higher level security servers where it can be tracked down and isolated.

Keywords Advanced metering infrastructure (AMI) · Petri net · Intrusion · Security

1 Introduction

During the last few decades, there have been quite a few innovative approaches to improve the performance of electrical power grid and make it more responsive to the needs of the end users. The areas that have been in focus include different perspectives like poor, outdated infrastructure, primitive power generation, distribution, transmission and consumption technology. These challenges are burdens on the economic infrastructure of any country. Eventually, smart grid has been proposed as the

D. Ghosh · N. Chaki (✉)
University of Calcutta, Kolkata, India
e-mail: nabendu@ieee.org

D. Ghosh
e-mail: debrajrock@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_39,

solution to get rid of these problems. Smart grid is the way of interconnecting the electrical power system to the highly sophisticated computer technology for ensuring smart infrastructures for the four aspects of power generation, transmission, distribution and consumption [1]. Smart grid replaces the traditional electrical meters with smart meters equipped with advanced metering infrastructure (AMI). However, smart grid is vulnerable to both physical and cyber threats. Physical intrusion is done by stealing, corrupting those costly electrical instruments. Sometimes, these result in power outages or black outs. Meter tampering is also treated as one way of intrusion. In the era of smart grid, the power grid is connected via computer network which in turn opens a new vulnerability level as computer networks are come under cyber-crime issues. Smart meters are connected to the consumers under any local substation. Thus, intrusion on smart meters has become much easier as the intruder can tamper the meter reading on his/her own smart meter just by corrupting the meter data if he/she has enough knowledge of cyber-crime. Corruption of instruments may be done due to catastrophic failures also. Thus an outstanding defense mechanism is the call of the hour for the security of smart grids.

In this paper, a new hierarchical color net model has been proposed for power transmission and distribution from the main power station to the consumers. This model has been used to track down the intrusion at consumer level as the smart meter tampering made by the intruders and carry this security information (affected smart meter ids) to the higher level security servers.

2 Related Work

In [2], bandwidth allocation of mobile nodes in different zones in a mobile ad-hoc network is considered. A two tire hierarchical colored PN model is proposed. Different zones are represented by different colors in this mobile ad-hoc network. The concept of hierarchy is different from other hierarchical structures. The tokens in each place of the higher level Petri net represent their corresponding sub PN in the lower level. In [3], the formal definition of hierarchical colored Petri net is given (HCPN). Colored Petri Net is reused for modeling several problems in this work. The selection property is not supported by RCPN perfectly. Paper [4] is based on identifying the faults in data transmission and faults in both power and communication in a smart grid. Smart grid is becoming the target of the coordinated attack. A novel net model is invented to model these coordinated attacks in [5]. Some key issues are discussed as the coordinated attacks, physical threats, cyber threats, and coordinated threats. In [6], a generalized stochastic Petri net model is considered for power generation from both renewable and nonrenewable energy sources. In [7], a Hierarchical Synchronized Multimedia integration Language (H-SMIL) Net model is developed. This model is an extended version of the SMIL net. In [8], a dynamic forensics system (DFS) is introduced by using Hierarchical Fuzzy Colored Petri Nets (HFCPNs). This system is designed to detect suspicious or malicious packets in the network. The model is divided into different levels for different specific tasks. This model is comparatively simple. Assembly and control planning is represented by Petri Nets in [9]. The use of electrical energy is made reliable using smart grid in [10]. This is done based on the power demand task scheduling policy as defined by the controller to reduce the grid operational cost. If simple fuzzy Petri nets are used in complex knowledge systems, the complexity of the system increases. In order to reduce the complexity, hierarchical fuzzy Petri net (HFPN) is used in [11]. A multistage attack model is introduced based on hierarchical colored Petri net for dealing with network security in [12].

3 Proposed Petri Net Based Model for Smart Grid

Petri Net (PN) is a tool based on bipartite graphical structure which models dynamic systems. A PN consists of two types of nodes places and transitions. A place represents the state of the modeled system and is denoted by circle whereas a transition, denoted by a bar, represents the computational activities carried out by the system when it changes its state. A marking function is attached to each place which maps each place to a value representing number of tokens in it.

3.1 Model for Power Distribution

The proposed model is based on four tire architecture layer 1, layer 2, layer 3, layer 4. The Petri net corresponds to layer 1 is shown in the Fig. 1a. In layer 1, there is a place Pr. It is assumed that place Pr represents the main power station where 2 MW power is generated from 11 KV. The places P1, P2, P3….P n denote the main substations that are directly connected to main power station Pr through transmission lines. From place Pr, power is distributed to the places P1, P2… P n through the transition Tr. The transition Tr acts as a step up transformer which converts the 2 MW power from 11 KV to 2 MW power from 132 KV. The conversion to high voltage is done due to get rid of power loss. After the transformer operation, power is distributed to the individual main substations. At each substation, the power is 2 MW from 132 KV. There is a step down transformer connected to each main substation. This transformer transforms the power of each main substation from 2 MW from 132 KV to 2 MW from 11 KV. Step down transformers connected to each main substations are denoted by the transitions T1’, T2’,…,Tn’. In the second layer, the area covered by the main substation P1 is shown in Fig. 1b. The place P1 represents the main substation with 2 MW power from 11 kV because the voltage was stepped down from 132 to 11 KV in the layer 1 by the step down transformer (T1’). The power of the main substation P1 is 2 MW. That means 2000 KW. This power is divided into four substations each having 500 KW from 11 KV. The area covered by the substations having power 500 KW from 11 KV is denoted by four places P1.1, P1.2, P1.3, and P1.4. The power is distributed from substation P1 to all four transformers by the transition T1. There

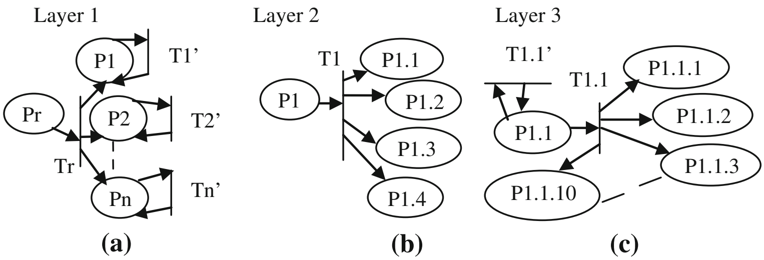


Fig. 1 **a** Layer 1 net model for main power station. **b** Layer 2 net model for main substation1. **c** Layer 3 net model for the area covered by substation 1 of main substation 1

are sub Petri nets for all rest of the substations P2, P3… Pn in layer 2. Here, only the sub Petri net for P1 is shown. As the model is based on hierarchical structure, the sub Petri net of this layer 2 is connected to its counter part of layer 1 by matching the identical places. In this case, all n main substations P1,…,Pn in the layer 1 represents their corresponding sub Petri nets which represent the area covered by each main substation. All individual sub Petri nets of layer 2 are connected to their corresponding main substations in layer 1 by matching the identical places. The sub Petri net of layer 2 is connected to the station P1 in layer 1 by matching the two identical places P1 both in layer 1 and layer 2. In the third layer, area covered by substation P1.1 is shown. In Fig. 1c, substation P1.1 supplies 500 MW power from 11 KV. This power is distributed to the sub areas covered by the utility poles. The power from the substation P1.1 is distributed to 10 utility poles under the area of this substation P1.1. Power at each utility pole is 50 KW. These utility poles under the area covered by P1.1 are represented by the places P1.1.1, P1.1.2, P1.1.3… P1.1.10. Before distribution of power from P1.1 substation, the voltage is stepped down from 11 KV to 415 V. For this operation, there is a transformer T1.1' is connected to P1.1. T1.1' steps down the voltage from 11 KV to 415 V. After stepping down of the voltage in substation P1.1, the power is distributed to each utility pole under the area of that substation. Thus, power at each utility pole is 50 KW from 415 V. The distribution of power from P1.1 to the utility poles is done by the transition T1.1. Like substation P1.1, there are also other substations P1.2,.., P1.4. They have their own utility poles in this layer 3. The four substations are connected to their corresponding covered areas represented by their individual sub Petri nets in layer 3. Like layer 1 and layer 2 the two layers, layer 2 and layer 3 are also connected by matching the identical places. In this scenario, the sub Petri net under P1.1 in layer 3 is connected to the substation P1.1 in layer 2 by matching both the identical places P1.1 of layer 2 and layer 3. In the fourth layer, the area covered by the utility poles is shown.

The area under one utility pole under the substation P1.1 is shown in this Fig. 2. This utility pole is represented by the place P1.1.1. At this utility pole P1.1.1, power is 50 KW from 415 V. There are five consumers are connected to each utility pole. Each consumer consumes 10 KW power from 415 V. It has been assumed all consumers consume same amount of current for simplicity. In this case, each consumer

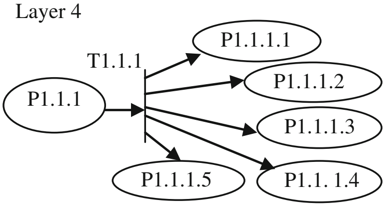


Fig. 2 Layer 4 net model for the area covered by pole1 of substation 1 of main substation 1

needs 10 KW power from 415 V. The five different consumers under the utility pole P1.1.1 are denoted by the places P1.1.1.1, P1.1.1.2, P1.1.1.3, P1.1.1.4, and P1.1.1.5. Distribution of power from the utility pole P1.1.1 to all consumers under this utility pole is done by the transition T1.1.1. Like P1.1.1 all other nine utility poles under the substation P1.1 also have their own five consumers. Like layer 2 and layer 3, the sub Petri nets of layer 3 are also connected to their corresponding sub ordinates. Each utility pole denoted by the places P1.1.1, P1.1.2.... are representing a sub Petri net as shown in layer 4. The utility poles of the layer 3 are connected to their corresponding sub Petri nets of layer 4 by matching the identical places of both the two layers layer 3 and layer 4. In this scenario, the corresponding sub Petri net of utility pole P1.1.1 of layer 3 is shown in layer 4. This sub Petri net covers the consumers under the utility pole P1.1.1. The utility pole P1.1.1 of layer 3 is matched with its sub Petri net of layer 4 by matching both two identical places P1.1.1 of layer 3 and layer 4.

3.2 *Model for Intrusion Detection*

In terms of intrusion detection, smart meter tampering detection is considered only. The Petri net related to meter tampering is also of hierarchical structure. Two levels of hierarchy are there. Not to make the model too complex and for reducing state space problem, layer 4 is not taken into consideration. Modeling has been started from layer 3. In layer 3, data from individual consumers are collected under area of each utility pole. In layer 3, GPS systems are used to collect meter readings for all consumers under each utility pole. Meter reading is available from the consumers' smart meter also from the utility pole connected to that consumer. So, the GPS system collects reading for each consumer not only from consumer's smart meter but also from the utility pole connected to that consumer. Thus, there are two places PC and PM corresponding to each utility pole. Meter readings collected from all consumers' smart meter and from the utility pole corresponding to all consumers' are stored in these two places respectively. This two meter reading information are compared to check whether they are identical or not. If the information is matched for an individual consumer then, there is no intrusion. If the information corresponding to a consumer is not matched, then there is an intrusion in the smart meter of that consumer. The information after matching of all consumers' under each utility pole is sent to corresponding substation connected to the utility poles. Each substation in layer 3 is represented by a place. In Fig. 3a, the substation of 500 KW from 415 V is represented by the place PM1.1. The places PM1.1.1, PC1.1.1… PM1.1.10, PC1.1.10 are representing the ten utility poles connected to this substation PM1.1. In PM1.1.1, meter reading of all consumers under utility pole 1 are collected from that utility pole. In PC1.1.1, meter reading for all consumers collected from their smart meter.

The matching of the meter readings for all individual consumers under pole 1 is done in transition TM1.1.1. All the meter readings after matching are sent to the substation PM1.1 by TM1.1.1. After collecting data after matching for all consumers, the matched data are sent to the main substation of 2 MW from 11 KV in layer 2. The hierarchy in this structure is only between layer 2 and layer 3. In layer 2, each of four substations is coming up with matched data of all utility poles connected to them. The matched data are transferred to main substation. From this main substation, all intrusion detection related issues are considered. This main substation finds out the intruders from all consumers connected to it by collecting data through the substations. Then, it takes actions against all of the intruders according to law. In Fig. 3b, only the main substation PM1 is considered. Under PM1, four substations are there PM1.1… PM1.4. The Petri net of layer 2 is connected to their corresponding sub Petri nets of layer 3 by matching identical places. In this case, the substation represented by the place PM1.1 of layer 2 is connected to the sub Petri net under this substation in layer 3 by matching the two identical places PM1.1 of both layers. The collected data after matching of all individual substations PM1.1… PM1.4 is sent to the main substation PM1. Only the matched data are sent to the substation PM1. Data is sent from all the substations PM1.1 to PM1.4 by the transitions TM1.1… TM1.4 connected to all individual substations PM1.1… PM 1.4. Due to limited space, only two places are assumed to be connected to each utility pole in layer 3. In Fig. 4a, the Petri net corresponding to pole 1.1.1 is shown. The place PM1.1.1 contains meter reading information collected from the utility pole 1.1.1 of all consumers under this pole. Each yellow token represents the information collected from the pole corresponding to the consumers in place PM1.1.1.

There are five consumers under each utility pole, so there are five yellow tokens in the place PM1.1.1. Each token contains ID of each consumer. The value on each token helps in identification of each consumers as, yellow token with value 1.1.1.1

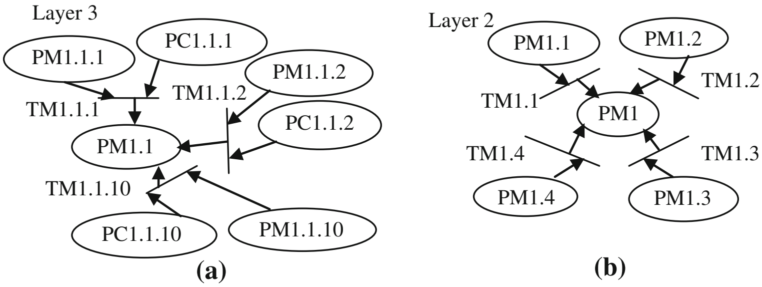


Fig. 3 **a** Layer 3 model for intrusion detection. **b** Layer 2 model for intrusion detection

represents the meter reading information of the first consumer collected from the utility pole 1.1.1. The meter reading information regarding to all consumers collected from their smart meter are stored in the place PC1.1.1. Each pink token in the place PC1.1.1 represents meter reading information regarding to each consumer collected from their smart meter corresponding to each consumer. Like yellow tokens, there are also five pink tokens regarding to five consumers in the place PC1.1.1 under the pole 1.1.1. Now, the transition TM1.1.1 is used to match the collected information for each consumer both from pole 1.1.1 and also from the consumer's smart meter. So, it is enabled whenever, there is yellow token in place PM1.1.1 and a pink token in PC1.1.1 regarding to a same consumer. After the firing of the transition TM1.1.1, the matched data regarding to meter reading along with the unmatched data is sent to the substation represented by the place PM1.1. Figure 4a shows the initial marking of the Petri net. Figure 4b shows the net after firing of the transition TM1.1.1. In place PM1.1, the green tokens represent the matched data of individual consumer and red tokens represent the unmatched data of individual consumers. The identification of all the consumers is also shown by value of the individual tokens. If the red token represents the consumer 1.1.1.1 then the red token represents that there is a mismatch in the meter reading of the consumer 1.1.1.1. The green tokens along with their levels represent the other consumers meter reading information is safe and sound. For simplicity, only the scenario of the pole1.1.1 is considered. But, this scenario will be repeated for all other consumers under other nine poles under the substation PM1.1. The transition TM1.1.1 fires whenever it is enabled. Thus the transitions of all the utility poles of all substations are immediate transitions. Figure 5a shows the scenario of layer 2. The Petri net under the area of the main substation 1 is shown. The four substations 1.1… 1.4 under this main substation come up with there meter reading information after matching of all consumers under the area of each substation. This figure shows the net before firing of any transition and Fig. 5b shows the model after firing of all the transitions. It is evident from the figures that after firing of all transitions, the red colored tokens are passed to the substation PM1. Thus, we can notice that intrusion is done in the smart meter of the four consumers.

Figure 5a shows the scenario of layer 2. The Petri net under the area of the main substation 1 is shown. The four substations 1.1… 1.4 under this main substation come up with there meter reading information after matching of all consumers under the area of each substation. This figure shows the net before firing of any transition

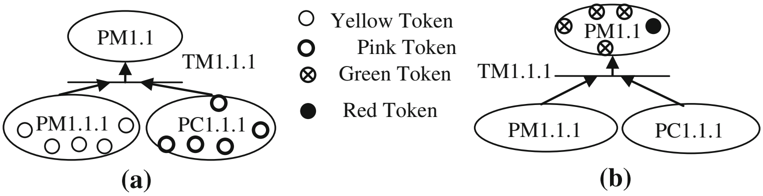


Fig. 4 **a** CPN model before the firing of the transition TM1.1.1. **b** CPN model after the firing of the transition TM1.1.1

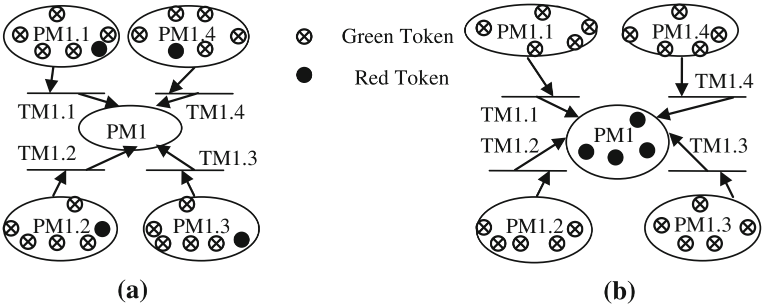


Fig. 5 CPN model **a** before firing of the transitions, **b** after firing of the transitions

and Fig. 5b shows the model after firing of all the transitions. It is evident from the figures that after firing of all transitions, the red colored tokens are passed to the substation PM1. Thus, we can notice that intrusion is done in the smart meter of the four consumers. We can recognize those consumers from the values of the red tokens. Here all the four transitions fire whenever they are enabled. So, these transitions are also immediate transitions.

4 Conclusion

This paper models the smart grid architecture using a hierarchical color net model. The model presented in Sect. 3.1 represents the power transmission and distribution network from the main power station to the consumers. Besides, it is described in Sect. 3.2 how the proposed model may be used to track the intrusion at consumer level. The work may be extended to ease the process of security monitoring and fault tracking by designing hierarchical control panels to be devised in different nodes in smart grid networks. In future, it has been planned to simulate the proposed model to verify and explore the potential of the proposed model.

References

1. Massoud BF (2005) Wollenberg: toward a smart grid: power delivery for the 21st century. IEEE Power Energy Mag 3(5):34–41
2. Dasgupta M, Chaudhury S, Chaki N (2012) A hierarchical CPN model for mobility analysis in zone based MANET. In proceedings of the fourth international conference on wireless, mobile network & applications, New Delhi, 25–27, May 2012. ISBN: 978-3-642-30110-0
3. Lee NH, Hong JE, Cha SD, Bae DH (1998) Towards reusable colored petri nets. In: proceedings of Int'l symposium on software engineering for parallel and distributed systems, p 223–229
4. Calderaro V, Hadjicostis CN, Piccolo A, Siano P (2011) Failure identification in smart grids based on petri net modeling. IEEE Trans Ind Electron 58(10):4613–4623

5. Chen TM, Aarnoutse JCS, Buford J (2011) Petri net modeling of cyber-physical attacks on smart grid. IEEE Trans Smart Grid 2(4):741–749
6. Dey A, Chaki N, Sanyal S (2011) Modeling smart grid using generalized stochastic petri net. J Convergence Inf Technol 6(11):104–114. ISSN: 1975–9320
7. Bouyakoub S, Belkheir A (2008) H-SMIL-Net: a hierarchical petri net model for SMIL documents. In: proceedings of the 10th IEEE international conference on computer modeling and, simulation, p 106–111
8. Chen HC, Sun JZ, Wu ZD (2010) Dynamic forensics system with intrusion tolerance based on hierarchical coloured petri-nets. In IEEE Int Conf Brd Wir Comp 660–665
9. Thomas JP, Nissanke N, Baker KD (1996) A hierarchical petri net framework for the representation and analysis of assembly. IEEE Trans Robot Autom 12(2):268–279
10. Koutsopoulos I, Tassiulas L (2010) Control and optimization meet the smart power grid: scheduling of power demands for optimal energy management. In proceedings of the Int'l conference on energy-efficient computing and networking, p 41–50
11. Pan H, Sun J (2007) Complex knowledge system modeling based on hierarchical fuzzy petri net. In: proceedings of IEEE/WIC/ACM international conferences on web intelligence and intelligent agent technology, p 31–34
12. Wu R, Li W, Huang H (2008) An attack modeling based on hierarchical colored petri nets. In: proceedings of IEEE international conference on computer and, electrical engineering, p 918–921

Part VI
The Fourth International Conference on Networks & Communications (NETCOM-2012): Network Security, Trust and Privacy

A Comprehensive Study on Two-factor Authentication with One Time Passwords

Kumar Abhishek, Sahana Roshan, Abhay Kumar and Rajeev Ranjan

Abstract Multifactor Authentication is a term used to describe security mechanisms that make use of two or more distinct and different category of authentication mechanisms to increase the protection for valid authentication. Multifactor Authentication that uses two levels of security can be used to obtain Two-factor Authentication. Introducing One Time Passwords (OTPs) to the existing authentication scheme can provide Two-factor authentication security. An OTP is a password that is valid for only one login session or transaction. OTPs avoid a number of shortcomings that are associated with traditional passwords. In this paper we do a survey on the different aspects of one time passwords and its impact on Two-factor Authentication.

Keywords Single factor authentication · Multifactor authentication · One time passwords · OTP

K. Abhishek (✉)
Department of Computer Science and Engineering, NIT Patna,
Patna 800005, India
e-mail: kumar.abhishek@nitp.ac.in

S. Roshan
Department of IT, Al Musanna College of Technology, Muscat, Oman
e-mail: sahana@act.edu

A. Kumar · R. Ranjan
Department of Information Technology, NIT Patna, Patna 800005, India
e-mail: abhay.kumar@nitp.ac.in

R. Ranjan
e-mail: rrnitp@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_40,

1 Introduction

An authentication system is how you identify yourself to the computer. The goal behind an authentication system is to verify that the user is actually who they say they are [1]. There are three major types of Authentication commonly used today. These authentications are based on:

- What you know (Knowledge Factors)
- Passwords, Personal Identification Numbers
- What you have (Ownership Factors)
- Keys, OTP, Tokens, Smart cards
- What you are (Inherence Factors)
- Finger Prints, Retinal Scans, DNA [2, 3]

Whenever one of the above schemes is implemented individually, it is referred to as Single Factor Authentication (SFA). In real time, SFA is most commonly implemented in a very traditional security process that requires a username and password before granting access to the user. SFA security relies on the alertness of the user, who should take additional precautions, for example by creating strong passwords and ensuring that no one accesses it.

Many problems exist within the world of single factor authentication. The first name of username/password combination, the username, may seem non-threatening in a security sense. However, in a single-factor authentication site, knowing the username or even the current naming convention of the usernames within an organization already gives the potential hacker 50 % of the information to gain access to vital information [4].

The features of this traditional (Username-Password) scheme can be listed as:

- Easy to implement
- Requires no special equipment
- Easy to forget
- Can be susceptible to shoulder surfing
- Security based on password strength
- Lack of identity check
- Cost of support increases

Long passwords are good, but consider the following:

User 1 password = badburger6
User 2 password = txyzuqne

User 1 password is made up of 2 words and one number, assuming 20,000 easy to remember common words in the English language, strength is 20K * 20K * 10 = 4 billion or in terms of cryptographic strength, a 32 bit key. User 2 password is eight characters randomly generated therefore strength is 26 to the power of 8 = 208 billion combinations or in terms or cryptography strength, a 38 bit key.

The User 2 password is stronger and surpasses the strength of user 1. However, to gain this sort of strength usually requires the user to have a photographic memory or to write this password down [5]. For applications that require greater security, it may be advisable to implement more than one type of the above mentioned schemes. The implementation thus gets termed as Multifactor authentication.

One problem with implementing multifactor authentication generally is the lack of understanding of "true" multifactor authentication. Supplying a username and password (Both being Knowledge factors) is single factor authentication despite being multiple pieces of distinct information. Supplying additional information in the form of answers to challenge questions (Again Knowledge factor) is still single-factor authentication. Adding a visual image for identification would still be single-factor authentication. An example of true Multi-factor Authentication is requiring the user to insert something (Ownership factor) or requiring a valid fingerprint via biometric reader (Inherence factor) [6, 7].

The pros of implementing Two-factor Authentication can be listed as follows:

- Reducing the Window of Opportunity
  The window of opportunity is wide and open with static passwords. A second factor of authentication can narrow down this window of opportunity and render any collected data useless.
- Eliminating Passive Attacks
  Implementing a two-factor authentication method, which minimizes the window of opportunity, can eliminate passive attacks as the stolen credentials are only valid for a short period of time.
- Limited validity for the data obtained
  Another benefit of two-factor authentication is that stolen data is only valid for single use and cannot be used for repeated access.
- Mitigating the Risk of Active Attacks
  Two-factor authentication with signing capability can prevent the attacker from achieving any financial benefit, since the transaction amounts and destinations can be digitally signed and will be of no use to the attacker.
- Increasing the Cost to Implement Fraud
  Implementing two-factor authentication will make it harder for fraudsters as they will have to shift to more expensive active attacks. Two-factor also reduces their gain as it shrinks the window of opportunity [8].

Multi-factor authentications products make are a need for the day today. A solution is required that should be easy to use and administer and when combined with username/password scheme should provide reasonable security. Costs are the biggest issues when it comes to multifactor authentication, but there are a variety of options at varying price points [9].

One form that can be used for Two-factor authentication along with the traditional username-password scheme is the concept of One Time Passwords (Fig. 1). A One-Time Password (OTP) is valid for only one login session [10]. A traditional, static password is usually only changed when necessary: either when it has expired or when the user has forgotten it and needs to reset it. Because passwords are cached on computer hard drives and stored on servers, they are susceptible to cracking. Unlike

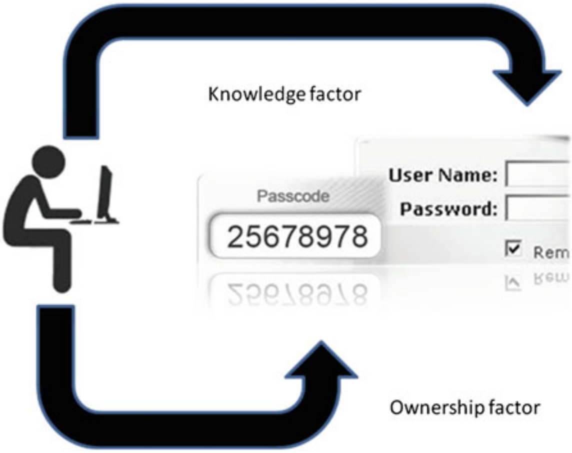


Fig. 1 An example for Two-factor Authentication

a static password, a one-time password changes each time the user logs in. A one-time password system uses a different password every time you want to authenticate yourself. Each password is used only once; thus, the term “one-time”.

The most important shortcoming that is addressed by OTPs is that, in contrast to static passwords, they are not vulnerable to replay attacks [11]. This means that, if a potential intruder manages to record an OTP that was already used to log into a service or to conduct a transaction; he or she will not be able to abuse it since it will be no longer valid. On the downside, OTPs are difficult for human beings to memorize. Therefore they require additional technology in order to work.

2 Types of OTP

OTP generation algorithms typically make use of randomness. This is necessary because otherwise it would be easy to predict future OTPs from observing previous ones. Concrete OTP algorithms vary greatly in their details. Various approaches for the generation of OTPs are:

- Based on time-synchronization between the authentication server and the client providing the password (OTPs are valid only for a short period of time). A time-synchronized OTP is usually related to a piece of hardware called a security token (e.g., each user is given a personal token that generates a one-time password). Inside the token is an accurate clock that has been synchronized with the clock on the proprietary authentication server. On these OTP systems, time is an important part of the password algorithm since the generation of new passwords is based on the current time rather than, or in addition to, the previous password or a secret key. This token may be a proprietary device, or a mobile phone or similar mobile device which runs software that is proprietary, freeware, or open-source [12].

- Using a mathematical algorithm to generate a new password based on the previous password (OTPs are, effectively a chain and must be used in a predefined order) [13].
- Using a mathematical algorithm where the new password is based on a challenge (e.g., a random number chosen by the authentication server or transaction details) and/or a counter. Each new OTP may be created from the past OTPs used. An example of this type of algorithm, credited to Leslie Lamport, uses a one-way function (call it f). The one-time password system works by starting with an initial seed s, then generating passwords f(s), f(f(s)), f(f(f(s))), .. as many times as necessary. If an indefinite series of passwords is wanted, a new seed value can be chosen after the set for s is exhausted. Each password is then dispensed in reverse, with f(f(...f(s))...) first, to f(s). If an intruder happens to see a one-time password, he may have access for one time period or login, but it becomes useless once that period expires. To get the next password in the series from the previous passwords, one needs to find a way of calculating the inverse function $f - 1$. Since f was chosen to be one-way, this is extremely difficult to do. If f is a cryptographic hash function, which is generally the case, it is (so far as is known) a computationally infeasible task. In some mathematical algorithm schemes, it is possible for the user to provide the server with a static key for use as an encryption key, by only sending a one-time password [14].
- The use of challenge-response one-time passwords will require a user to provide a response to a challenge. For example, this can be done by inputting the value that the token has generated into the token itself. To avoid duplicates, an additional counter is usually involved, so if one happens to get the same challenge twice, this still results in different one-time passwords. However, the computation does not usually involve the previous one-time password; i.e. usually this or another algorithm is used, rather than using both algorithms [14].
- In another flavor of the challenge based OTP, The authentication server displays a challenge (a random number) to the user when he attempts to authenticate. When the user enters the challenge number into the token, it executes a special algorithm to generate a password. Because the authentication server has this same algorithm, it can also generate the password. Thus the authentication of the user can be completed if the passwords match.

There are also different ways to make the user aware of the next OTP to use. Some systems use special electronic tokens that the user carries and that generate OTPs and show them using a small display. Other systems consist of software that runs on the user's mobile phone. Yet other systems generate OTPs on the server-side and send them to the user using an out-of-band channel such as SMS messaging. Finally, in some systems, OTPs are printed on paper that the user is required to carry with them.

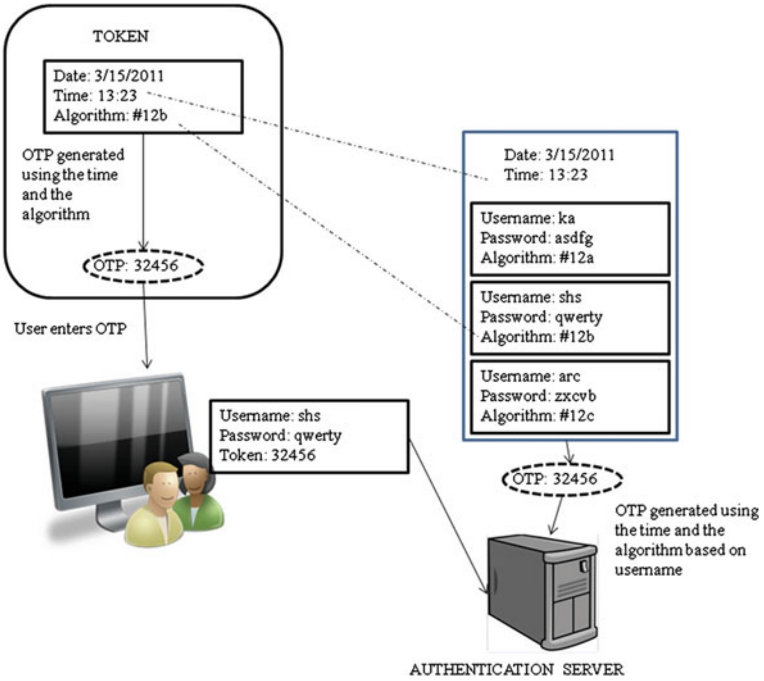


Fig. 2 Time -synchronized OTP

3 Advantages of OTP

One-time password systems can be easy to deploy and may not require any special software to be installed on the customer's computer. Some systems use one-time passwords generated on a hardware device that is communicated directly to the computer, say through a USB port. This option requires software to be installed (Fig. 2).

One-time password systems are generally acceptable to customers, due to their similarity to password systems. One-time password clock-based devices and challenge/response systems can be used across multiple systems (whereas counter-based solutions cannot without complicated re-synchronization). It is necessary that these are trusted systems, as each has the capability to impersonate the customer to the others. In practice, clock-based systems may also require time synchronization to work effectively.

With hardware one-time password devices and printed lists, the customer is likely to notice the loss if they are stolen [15, 16].

4 Disadvantages of OTP

The verifier will need special software and/or hardware. Protected storage and management of the base secrets is required.

A disadvantage with clock-based one-time passwords used across multiple systems is that there is a window of exposure: when a one-time password is used it can be used with any of the other systems if an attacker obtains it. Shorter windows reduce the scope of such attacks. Also, these attacks may be countered by protecting the communication channel. Most hardware one-time password devices do not provide the same level of tamper resistance, and thus protection for the base secret, as hardware tokens do. This may change in the future as the hardware one-time password device market matures.

Systems based on shared printed tables, sometimes called bingo cards, have the same problems as written-down passwords: they may be copied or discovered and used without the customer's knowledge. Loss of the authentication key itself is a much more severe breach of security than the loss of any single one-time password. Shared tables exist that conceal the numbers under a coating, called scratchy cards, with the customer removing the coating to reveal each one-time password. These cards defend against copying attacks. They may still be stolen and used, although the customer would be expected to notice the loss of their card.

With authentication key sharing, the extent of the problem here would relate to how easy it is to copy. If copying is easy, then the customer can share their authentication key without losing the ability to authenticate. If copying is not feasible, then this may deter customers from sharing their authentication key, as they must also give up their ability to authenticate [16].

5 Implementation and Security Aspects of One-Time Passwords

The main security property that protocols employing one-time passwords should achieve is: strong mutual authentication based on knowledge of one-time passwords. The motivation for using one-time passwords is that the compromise of one password should not affect the security of sessions involving another password. The one-time password serves to mutually authenticate the client and the server; there are no other long-term values like public keys or certificates. Authentication is based on knowledge of the shared password. A one-time password protocol should give secure mutual authentication for the current session even if other one-time passwords have been revealed. In addition to mutually authenticating two parties to each other, a protocol should also output a session key that can be used to encrypt and protect the integrity of future communications between those two parties. This is a common feature required of many secure communication protocols [17].

Table 1 Different aspects of OTP [8]

Protection against passive attacks	High
Protection against active attacks	Low to medium
Initial cost involved	Medium
Support and usage costs	Low to none
Ease of use	Medium
Portability	High
Special software for client required	No

The table provides the general details of OTP from a security, cost and usability perspective. The generalization may differ affected by the difference in implementation (Table 1).

The success of the OTP system to protect host systems is dependent on the non-invert ability of the secure hash functions used. If a server supports multiple hash algorithms, it is only as secure as the weakest algorithm.

One-time passwords in general can withstand replay, eavesdropper, key logger and shoulder-surfing attacks; because once a one-time password is used it cannot be used again. One-time passwords used across multiple systems cannot completely mitigate against these attacks without further protection measures being in place. Using communication channel protections mitigates session hijacking attacks. Other attacks are not mitigated by one-time passwords themselves. Systems should employ further protections for the communication channel. The scope of customer fraud attacks would depend on the actual product (primarily this relates to the easy of copying and tamper resistance features). An important distinction with passwords is that a phishing attack only gains a single one-time password, which greatly decreases the scope of these attacks when compared to passwords. Phishing, a method commonly used by fraudsters in recent years, is an example of a passive attack. The term phishing comes from the analogy that hackers fish for user credentials on the Internet by putting bait in front of users to lure them into a trap. The bait is usually an innocent looking email with a cover story to convince users to visit a counterfeit website, to divulge credentials and personal information. To date, phishing scams have been passive, largely due to the fact that harvested data can be used at a later time [8]. This may change in the future as two-factor authentication becomes more prevalent and the window of opportunity for exploiting harvested data shrinks. In late 2005 customers of a Swedish bank were tricked into giving up their one-time passwords. In 2006 this type of attack was used on customers of a US bank.

Although OTPs are in some ways more secure than a memorized password, users of OTP systems are still vulnerable to man-in-the-middle attacks. Man in the middle (MIM) is a term borrowed from cryptography where an attacker gains access to the secret key used for encrypting data between a sender and a receiver. The attacker can then eavesdrop between the two parties while passing the information at the same time. In the context of credential theft, this is a counterfeit website that interacts with the user on behalf of the real site and passes the information behind the scenes between the two parties. MIM is an example of an active attack [8]. Some security

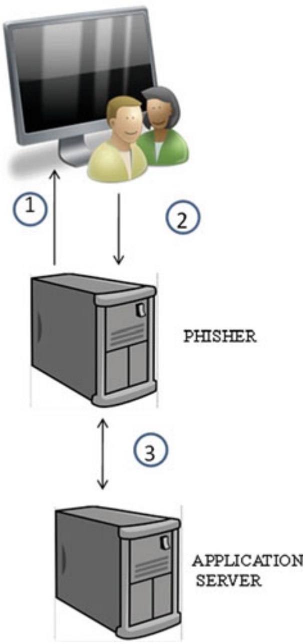


Fig. 3 Sample man-in-the-middle attack. *1*. Phisher sends legitimate looking request to log into familiar application. *2*. User recognizes URL and graphical interface and then enters password. *3*. Phisher authenticates using stolen credentials

analysts have used the hypothetical example of MIM attack to question the case for implementing two-factor authentication technology. While it is true that simply using a second factor authentication for user logins cannot prevent an MIM attack, a proper implementation of two-factor authentication across sensitive resources in tandem with complementary preventative technologies can neutralize or at least contain MIM attacks (Fig. 3).

OTPs should therefore not be disclosed to any third parties, and using an OTP as one layer in layered security is safer than using OTP alone; one way to implement layered security is to use an OTP in combination with a password that is memorized by the user (and never transmitted to the user, like OTPs often are).

The use of the OTP system only provides protection against passive eavesdropping/replay attacks. It does not provide for the privacy of transmitted data, and it does not provide protection against active attacks such as session hijacking that is known to be present in the current Internet. The use of IP Security (IPSec) is recommended to protect against TCP session hijacking [18].

In terms of costs, the cheapest OTP solutions are those that deliver OTPs on paper, and those that generate OTPs on a device that someone already owns. This is because these systems avoid the costs associated with (re-)issuing proprietary electronic tokens and the cost of SMS messaging [19].

For systems that rely on electronic tokens, algorithm-based OTP generators must cope with the situation where a token drifts out-of-sync with its server if the system requires the OTP to be entered on a deadline. This leads to an additional development cost. Time-synchronized systems, on the other hand, avoid this at the expense of having to maintain a clock in the electronic tokens (and an offset value to account for clock drift). Whether or not OTPs are time-synchronized is basically irrelevant for the degree of vulnerability, it but avoids a need to reenter passwords if the server is expecting the last or next code that the token should be having because the server and token have drifted out-of-sync.

Compared to most proprietary hardware tokens, so long as one already carries a phone or another mobile device in one's pocket, users of mobile devices don't need to carry and protect an extra item. In addition to reducing costs considerably, using a phone as a token offers the convenience that it is not necessary to deliver devices to each end-user. However, most proprietary tokens have tamper-proof features [19, 20].

6 Conclusion

There is no compromise with security. The basic need for any authentication mechanism is to establish a proof of identity to applications. Basic username/password mechanisms fail to do this. Strong Two-factor authentication using One Time Passwords mitigates against the threats in a non-real time environment. Most Companies have already implemented two-factor Authentication to secure their business assets and resources. Some Companies specializing in Two-factor Authentication include: ALADDIN, AUTHENEX, and RSA SECURITY, SECURE COMPUTING, VASCO and VERISIGN. OTPs all have their features and benefits, and the technology used is quite sophisticated and leaves little room for direct attacks. OTP tokens provide a reliable way to secure accounts. That's why it is used by firms engaged in sensitive industries including financial services. Innovative approaches are required to enhance One Time Passwords for protection against real-time phishing attacks and other emerging threats. There is a need to implement One Time Passwords which is protected by hashing or other methods and which expires before the attacker can recover it.

References

1. http://www.garudaindia.in/html/pdf/bootcamp2010/grid_security_igca.pdf
2. http://www.ffiec.gov/pdf/authentication_guidance.pdf
3. http://computer.yourdictionary.com/authentication
4. http://www.infosecwriters.com/text_resources/pdf/Two_Factor_Authentication.pdf
5. http://www.securenvoy.com/WhitePapers/white_paper_two_factor_authentication.pdf
6. http://blog.tevora.com/authentication/multifactor-authentication-2/
7. http://datariskgovernance.com/risk-assessment/multi-factor-authentication-in-banking/
8. Gpayments, Two-factor authentication: an essential guide in the fight against Internet fraud, February 2006

9. http://searchsecuritychannel.techtarget.com/feature/Does-the-customer-need-a-multi-factor-authentication-solution\
10. http://www.gemalto.com/brochures/download/ent_otp_secure_access.pdf
11. http://www.quuxlabs.com/blog/2010/09/paper-token-gutenbergs-version-of-one-time-passwords/
12. http://www.otptoken.com/
13. http://chain-one.blogspot.com/2010/11/one-time-password-history.html
14. http://www.ndkey.com.cn/otp_tech_introduction.html
15. Sangheethaa S, Swathika R and Sasirekha S (2008) Critical file access in wireless networks using multifactor authentication
16. http://www.e.govt.nz/standards/authentication/guidance-on-multi-factor-authentication/detailed-discussion-of-authentication-keys
17. Steinfeld R (2010) Information security and privacy. In: 15th Australasian Conference, ACISP 2010
18. Sen ÇAKIR, Fatih UÇAR Effectiveness of two factor authentication for preventing fraudulent transactions during session hijacking attacks on business, December 2007
19. http://www.enotes.com/topic/One-time_password
20. http://wn.com/one-time_password

ITRANS Encoded Marathi Literature Document Relevance Ranking for Natural Language Flexible Queries

V. M. Pathak and M. R. Joshi

Abstract Information Retrieval (IR) is an ever evolving concept of computer and communication applications. The emergence of mobile technology and its widespread service domains have given new dimensions to this as a research topic. Number of key players of search industry, are coming up with problems and their solutions with respect to SMS based Information systems and retrieval facilities. An extensive literature survey reveals that SMS based IR models for Indic languages are not available at large scale. In addition to this an universal transliteration encoding formats for Indic languages named ITRANS, has attracted us to resolve the problem of a suitable knowledgebase. With all these exploration we have formulated our research problem by combining both these features Viz. ITRANS encoded document corpus and SMS based Information Retrieval on this corpus. With reference of a few initiations of this emerging topic, we present an extended IR model for Marathi documents in ITRANS format. A collection of approximately hundred documents including Marathi songs, poems, abhangas and powada is incorporated as the document corpus for our experimental work. Several queries are formulated in ITRANS format and are used for the testing. The system developed by us ranks the documents based on the similarity scores applying Cosine Similarity measures. The documents in their ranked orders are checked by an expert for their relevance with respective queries. The feedback submitted by the expert is deployed to compute the relevance measure in terms of precision and recall. Finally the results are analyzed to correlate the similarity degree to the relevance of documents.

Keywords Information retrieval · ITRANS code · Vector space model · Cosine similarity · SMS based IR

V. M. Pathak (✉)
Department of MCA, IMR Jalgaon Affiliated to NMU, Jalgaon, India
e-mail: vm_pathak66@yahoo.co.in

M. R. Joshi
Department of Computer Science, North Maharashtra University, Jalgaon, India

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_41,

1 Introduction

Natural language query based information retrieval is the obvious part of information and communication technology. With emergence of mobile technology "SMS based Information Systems", has surfaced out of its service domains [1]. To extend our contribution in this field, we are looking for a suitable IR model in this context to deliver relevant information on mobiles as demanded by the user. We will be using the library domain as the backend information source in this problem. "Indian language literature in ITRANS encoded format can be made retrievable on the hand held device as and when asked by the user", is the core theme of our research problem.

Number of key players are coming up with solutions to the SMS based IR problems in various domains. Kopparappu et. al. for example deals with the access of yellow pages information using SMS [2]. An ontology based NLP engine is developed and experimented by them to process Natural Language SMS demanding the information related to yellow page databases. Pratima Mitra et. al. [3] handles the agriculture domain to drive the SMS based consultancy for rural based agriculture related queries and commodity booking services. This system is also designed to access related database using internet and local data connection as well. Many of such systems could be found for number of domains like education [4], agriculture [3], m-governance [5], m-banking and many more. These systems are either based on format free NLP queries or fixed format planned queries [1]. The need of NLP queries to access information as and when required from a knowledge base has impressed many researchers. The new concepts like SMS based Information Systems and Information Retrieval have emerged from this need [1, 6].

We have chosen Indic Language Literature Retrieval in the context of SMS based Information Retrieval Systems as our research domain [7–9]. This paper is an extension of the work that we have already presented [10] in this respect. As mentioned in this paper we are using collection of ITRANS Marathi Documents as the knowledge source of our experimental system. The Vector Space Model presented in this paper [10] is modified by applying TF-IDF formulation to compute the Cosine Similarity. We formulated our experiments to understand the correlation between the Cosine Similarity measure and the relevance of the document with the specified query.

The work presented in this paper includes the methodology used in the experimentation, the problem definition, the experimental model, analysis of the results in consecutive sections. The paper is concluded by highlighting the conclusions drawn out of the analysis and the future aspects of this topic.

2 Methodology

The main aim that we hypothetically want to achieve as part of the underlined research work is to answer the user's requested query by extracting the relevant contents out of the relevant documents. Before extracting the contents as per the user demand expressed in the query, we need to test the relevance of the documents based on the similarity measures. The similarity measures are computed based on term matching [11]. Basically the term matching is the “string matching”. We want to a add context based Similarity Measures in future. But initially we set our model based on the original concept of the Similarity Formulation that uses “String Matching”.

2.1 System Design

The system is designed to produce the Cosine Degree as the Similarity Scores for each document for the related query. Weights of query terms and document terms are computed using, TF-IDF formulation [6]. Based on the similarity scores, the system ranks the documents and delivers top ranked documents to the user for each query.

To mark the relevance of the ranked documents for the underline query, we engaged an impartial expert. Then we differentiated the lower bound of the similarity scores to make the respective ranked documents to get qualified for actual delivery to the user. This lower bound of the similarity score is termed as “Angular- degree level”. The list of documents qualifying “Angular- degree Level”, is finally delivered to the expert. The expert is asked to submit his/her opinion about the relevance of these documents for the concerned query. These “Angular-degree levels” are differentiated from a threshold degree up to a maximum level in successive runs. The same set of queries is applied in all runs. The experimental results are then tested by collecting expert’s feedback at each run. The results of randomly selected six queries are compared based on precision and recall. At the end of this paper, the outcome of the experimentation is presented and analyzed.

The most important part to develop an information retrieval model for the ITRANS Marathi literature is to collect a suitable size of the document base. As revealed in [10], we started collecting these documents available on http://www.sanskritdocuments.org. We have collected more than hundred documents from this source and processed them to construct a suitable document corpus. We collected number of queries from expert users who are familiar with ITRANS format. Then we devised our experiment based on the VSM using TF-IDF formulation for cosine degree measures [6].

2.2 Mathematical Model

We have formulated this model that uses a collection of documents D. The document vector and inverse document term vector are built over D. Similarly it acquires a query vector and inverse query term vector from the query base.

Let us define a query set Q = {q_1,q_2....q_n} and derive a matrix W_Q as in expression 1. Similarly let us define a document corpus D. D = { d_1, d_2,.....d_m }, where d_i is the ith document.

$$w_Q = \begin{bmatrix} w_{11}, w_{12} \ldots\ldots w_{1} \\ w_{21}, w_{22} \ldots\ldots w_{2l} \\ . \qquad\qquad . \\ w_{n1}, w_{n2} \ldots\ldots w_{nl} \end{bmatrix} \tag{1}$$

Where 1 = MaxNo{Terms{q_i □ Q, i = i..n}} and
w_{ij} = Weight calculated by expression 4.

$$w_Q = \begin{bmatrix} w_{11}, w_{12} \ldots\ldots w_{1p} \\ w_{21}, w_{22} \ldots\ldots w_{2p} \\ . \quad . \quad . \quad . \\ w_{m1}, w_{m2} \ldots\ldots w_{mp} \end{bmatrix} \tag{2}$$

Where p = MaxNo{Terms{d_i □ Q, i = i..n}} and
w_{ij} = Weight calculated by expression 5.

From this derive a matrix of weights of all terms in the document set as in expression 2. In the next step for all query q_i in Q, compute Similarity Measure applying the expression 3.

$$\text{Sim}\left(\frac{j}{i}\right) = \frac{\sum_{k=1}^{n} WkiDkj}{\sqrt{Wk^2}\sqrt{Dkj^2}} \tag{3}$$

Where Sim(j/i) is the Cosine Similarity of jth document with the ith query.
W_k is weight of kth query term of query q_i, computed by expression 4.
D_{kj} is the weight of that kth query term of q_i in jth document by applying expression 5.

$$W_i = t_{Qi}(\log_{10}(N/dt_{Qi})) \tag{4}$$

Where N is size of query set.

$$D_{ij} = t_{Dij}(\log_{10}(M/dt_{Dij})) \tag{5}$$

Where, M is the size of Document corpus. td_i is the term frequency of the same ith term in the document $d_{j,}$ and dtd_i is the number of occuerences of that ith term in the document set as a whole. M/dtd_i is termed as inverse document frequency of the term in document corpus.

3 Results and Analysis

As mentioned in previous section, we launched the consecutive experiments. In these experiments, iteratively size of document corpus and query set was increased. We started with the corpus size as 40 then 50, 60, 70 and 80 at final takes. At every iteration, the system produced a set of top ranked documents as d_1, d_2, $d_3 \ldots\ldots d_{10}$. At the specified rank position (i), document number (j) is assigned in the result. It indicates that the document d_j is relevant to the respective query with the respective rank number (i). The documents with the same similarity scores are ranked in increasing order of their positions in original corpus. Results based on the varied angular levels as mentioned in previous section, is presented to an impartial expert observer.

3.1 Sample Output

We present here Table 1 as the output of one of the iteration with five top ranked documents for two different queries with corpus size sixty. Where the first row of each query entry gives document numbers (between 1 and 60), at ranks 1–5 and second row of the same query entry displays the similarity score computed by TF-IDF formulation for respective documents. The column "Query" gives two examples of Marathi query in ITRANS format.

3.2 Precision and Recall

We applied these measures to test our system. For this we assigned one observer who has the understanding of ITRANS encoding. At each iteration the Angular-degree Level α is set to certain value. The expert is asked to indicate the relevance of the qualified documents as "YES" or "NO". This information is then used to compute Precision and Recall applying expressions 6 and 7 respectively [6].

We obtained the average precision and average recall, for each of the iterations. The average precision and average recall values are given the overall system performance. These results of precision and recall along with the varied α values and average number of documents retrieved in each iteration. This result is presented and

Table 1 Sample queries and document ranking

Run : Documents :: Doc # 1 to 60

Query #	Query	Rank #	1	2	3	4	5
Q1	aaj kuNitarii yaave giitaachi	Doc #	10	31	14	16	27
	gaayikaa koN	Score	0.639	0.452	0.452	0.452	0.452
Q2	airaNiichyaa devaa koNi	Doc #	12	18	22	39	45
	gayileya	Score	0.657	0.465	0.465	0.465	0.465

analyzed as in Chart 2.

$$\mathrm{P_i} = \frac{\text{Number of documents retrieved}}{\text{Number of relevant documents}} \tag{6}$$

$$\mathrm{R_i} = \frac{\text{Number of relevant documents retrieved}}{\text{Total number of relevant documents}} \tag{7}$$

The results of randomly selected six queries are analyzed as discussed in next section.

3.3 Result Analysis

The output of the initial runs with varied sizes of document corpus, are analyzed by us to gather following facts.

On the basis of the relevance remarks obtained for the various α values the "Precision and Recall" values are manipulated and redirected them in tabulated form in the excel sheet as in Table 2 . This table is analyzed graphically in Chart 1.

3.4 Graphical Analysis

We have accumulated following facts from Chart-1.

At $\alpha = 0.55$, 100 % relevant documents are covered in the qualified documents. After that recall value decreases saying the decrease in the performance to extract relevant documents.

The graph of precision values shows, that the relevance correlate at maximum at the similarity degree level 0.55. Above this level the precision has decreased as the relevant documents are outranked by the system. From the Chart-2 we understand as below.

Table 2 Result set

Query#	$\alpha = 0.35$ No. Doc.	Prec.	Recall
1	5.00	0.20	1.00
2	3.00	1.00	1.00
3	3.00	0.33	1.00
4	10.00	0.10	1.00
5	1.00	1.00	1.00
6	10.00	0.10	1.00

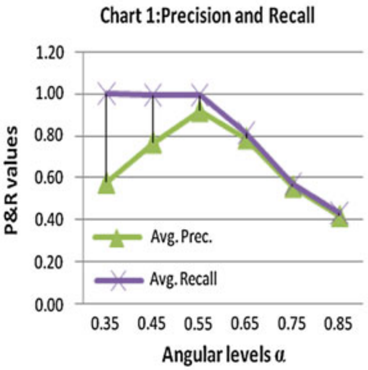


Chart 1:Precision and Recall

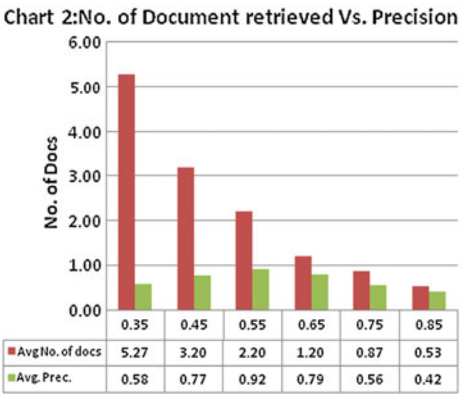


Chart 2:No. of Document retrieved Vs. Precision

	0.35	0.45	0.55	0.65	0.75	0.85
🟥 AvgNo. of docs	5.27	3.20	2.20	1.20	0.87	0.53
🟩 Avg. Prec.	0.58	0.77	0.92	0.79	0.56	0.42

The number of documents retrieved at angular levels from 0.35–0.85 has gradually reduced. It almost gives zero number of documents delivered to users at level more than 0.85. This means that even the relevant documents have scored less than 0.95.

4 Conclusion

This paper is about extending the TF-IDF based Cosine Similarity Model for ITRANS Indian Language literature retrieval. This model has been developed with the prime interest to obtain an information extraction model of SMS based information system for libraries. ITRANS encoded Marathi literature document set is formed for this purpose. The experiment was designed to check whether the Cosine Similarity is an applicable model in this problem. "The correlation of the Cosine Similarity to the relevance of the documents with the query", is investigated in this experimentation. From the explicit relevance feedback received by the expert users the results are analyzed. It shows that in all cases the highest scored documents are the most relevant documents for the corresponding query.

The important outcome of the research done in respect to the "SMS based ITRANS Marathi Literature retrieval system", is that the Cosine Similarity can be successfully applied to extract relevant documents from the document corpus with respect to a specified "Natural Language Query". For a very high Similarity Score as 0.95 the precision value decreases. From this we understood that the query term vector has to be refined so that the user's need of information access is manipulated properly. This would result in the relevant documents to be qualified at top ranks with higher Similarity Scores than 0.55. The relevance feedback mechanism is one of the solutions to refine the user query so that it matches with the relevant documents with very high scores.

As next part of our research we plan to explore and apply other IR models such as probability based IR model [12], support vector machine [13] in our problem.

We would also focus on Relevance Feedback Mechanism (RFM) [14], with respect to all these IR models to develop a suitable information extraction system for "ITRANS Marathi Literature Retrieval".

References

1. Joshi M, Pathak V (2011) A functional taxonomy of SMSbIR systems. In: 3rd International conference on electronics computer technology, Kanyakumari, 8–10 April, 2011, IEEE Xpolre, pp V6:166–V6:170 ISBN: 978-1-4244-8677-9
2. Kopparadu SK, Shrvastava A, Pande A (2007) SMS based natural language interface to yellow pages directory
3. Mitra P, Samajpati A, Sarkar T, Das PK (2006) An SMS based rural application for agricultural consultancy and commodity booking service
4. So S (2009) The development of a SMS-based teaching and learning system. Hong Kong Institute of Education, Hong Kong
5. Kushchu I, Kushchu H (2003) From E-government to M-government: facing the inevitable. In: The proceedings of european conference on E-government (ECEG 2003), Trinity College, Dublin, pp 1–13
6. LEE DL (1997) Document ranking and the vector space model. Hong Kong University of Science and Technology, HUII CHAUNG, Information Dimensions, KENT SEAMONS, Transarc, MARCH/APRIL 1997 0740–7459/97/$10.00© 1997, IEEE
7. Ran A, Lencevicius R (2007) Natural language query system for RDF repositories. In: Proceedings of the 7th international symposium on natural language processing, SNLP, pp 1–6
8. Goh TT, Li C (2009) SMS-based library catalogue system: a preliminary investigation of user acceptance. Victoria University of Wellington, Wellington, New Zealand, Liew school of information management
9. Rocchio JJ (1971) Relevance feedback in information retrieval. In: Salton G (ed) The SMART retrieval system: experiments in automatic document processing, Prentice Hall Inc, NJ, pp 313–323
10. Pathak V, Joshi M (2012) ITRANSed marathi literature retrieval using SMS based natural language query. Adv Comput Res 4(1):125–129
11. Soubbotin MM, Soubbotin SM (2002) Use of patterns for detection of answer strings: A systematic approach. In: Proceedings of TREC, vol 11. Citeseer, pp 134–143
12. Robertson SE (1977) The probability ranking principle in IR. J Doc 33(4):294–304
13. Drucker H, Shahrary B, Gibbon DC (2002) AT & T Research and Monmouth University, West Long Branch, NJ 07764 USA, Support vector machines : relevance feedback and information retrieval. Inf Process Manage 38:305–323
14. Salton G, Buckley C (1990) Improving retrieval performance by relevance feedback. J Am Soc Inform Sci 41:288–297

AB-OR: Improving the Efficiency in Onion Routing Using Attribute Based Cryptography

Nishant Doshi and Devesh Jinwala

Abstract Onion routing has been employed as a technique for ensuring anonymous communication over a public channel to ensure the integrity of the message as well as the privacy of the contending entities amidst untrusted routers. The technique ensures that the actual data as well as the meta-data in the packet are hidden from the prying intermediaries. It uses multiple layers of encryption of the packet for the purpose. However, exactly due to the same, the efficiency concerns in onion routing have to be addressed. In this paper, we propose and demonstrate the use of the Ciphertext Policy Attribute Based Encryption (CP-ABE) to improve the overhead in the existing onion routing schemes as well as to improve their overall security strength. Moreover, we also impart failure tolerance in case an Onion Router in the communication path breaks down. We formally call the proposed approach as AB-OR (Attribute Based Onion Routing).

Keywords Attribute · Attribute based cryptography · Network Security · Onion routing

1 Introduction

The rapid growth in internet communication has raised serious concerns about various security related issues. One of the vital security issues is the anonymity. Using the conventional techniques in Information Security practices, a sender and receiver can easily ensure the confidentiality and integrity of the data in the message. However, it

N. Doshi (✉) · D. Jinwala
Computer Engineering Department, Sardar Vallabhbhai National Institute of Technology, Surat, India
e-mail: doshinikki2004@gmail.com

D. Jinwala
e-mail: dcjinwala@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_42,

is equally vital to ensure that the anonymity of them both, as well as that of the path used in communication be protected from the prying eyes. The vulnerability to prying and eavesdropping increases manifold due to the multi-hop communication typically employed in the Internet. For example, merely by inspecting the *destination address* field in the packet, an adversary can easily gain knowledge about the recipient.

One of the ways of dealing with this issue is anonymity in routing. The technique of Onion Routing is an attempt in this direction. Proposed first by Goldschlag et al. in [1, 2] in 1999, Onion routing is based on employing multiple layers of encryption of a packet to be decrypted at each intermediate router known as an Onion Router (OR). The source uses the public keys of the intermediate routers with the top layer encrypted with the public key of the router immediately next to the source. The intermediate routers then use their own corresponding private keys to decrypt the packet and obtain the information about the next hop in the network. The packets thus routed and forwarded by each intermediate genuine node, eventually reach the destination; as in conventional networks. However, the advantage here is that if any one of the routers is compromised by an adversary, even then, the other components remain beyond the reach, because of being encrypted using a different public key.

From the basic paradigm followed in Onion Routing, it is evident that a sender is required to encrypt a message as many times as is the number of intermediate ORs. That is, if there are 100 ORs in between a sender and a receiver, then a sender is required to encrypt the original data packet 100 times, each with the public key of different ORs in-between. This exorbitantly increases the associated overhead at the sender. We exemplify this operation in the diagram shown in Fig. 1. In Fig. 1, the sender will select the path and based on the public keys of ORs in the path, encrypts the message and send to first OR. Here every OR is equipped with the pair of public and private key. The more details on working of the onion routing concept can be found in [3].

In addition, as per the existing onion routing schemes, an intermediate router is by default, required to know the identity of the next router in the path. However, if this condition is somehow eliminating, then the anonymity can be further increased.

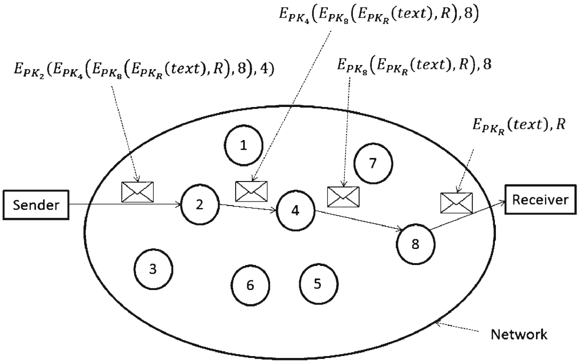

$E_{PK_4}(E_{PK_8}(E_{PK_R}(text), R), 8)$
$E_{PK_2}(E_{PK_4}(E_{PK_8}(E_{PK_R}(text), R), 8), 4)$
$E_{PK_8}(E_{PK_R}(text), R), 8$
$E_{PK_R}(text), R$
1
7
Sender
2
4
8
Receiver
3
6
5
Network

Fig. 1 Onion routing operation

Thirdly, the failure tolerance of the existing scheme is poor. That is, the failure of any one of the intermediate OR leads to the failure of the entire onion routing protocol.

In the proposed work, we attempt to overcome all of these limitations. We propose a protocol viz. *Attribute Based Onion Routing (AB-OR)* that employs Attribute Based Encryption for the purpose. In our scheme, a sender is required to encrypt the message only once irrespective of the number of intermediate ORs in the path. We are able to do this by designing a single policy for intermediate ORs. In addition, a potential replay attack that can occur in the default onion routing protocol can also be overcome in AB-OR by augmenting the attribute sets used for encryption with the *time* attribute. This would enable a data packet to be decrypted within a specific time frame only.

In addition, we also try to add failure tolerance in AB-OR. As compared to the default onion routing scheme in which the sender is required to have the knowledge of only one communication path and the intermediate routers therein; in AB-OR, we let the sender select any one of the path through which to forward the packet out of the multiple paths at its disposal. The assumption herein is that a sender with a priori knowledge of the probability of failure across all the paths to its disposal should select a path with the least probability of failure, thereby increasing its tolerance to failures.

One can see numerous attempts in the literature [4–27] that aim at overcoming the limitations of the onion routing protocol. However, while reviewing the related work, to the best of our knowledge, none of them attempt to address the issues that we focus on, in this paper. The notion of attribute based encryption can be understood from [28–34].

Organization of the paper: In Sect. 2, we discuss the preliminaries, which we use in the remaining part of the paper. In Sect. 3, we present the proposed approach. In Sect. 4, we give the analysis of the proposed approach, whereas conclude with the probable future work in the last section.

2 Preliminaries

2.1 Bilinear Group

Definition 1 (Bilinear map). Assume G1, G2 and G3 are three multiplicative cyclic group of some prime order p. A bilinear map $\mathrm{e} : \mathrm{G1} \times \mathrm{G2} \rightarrow \mathrm{G3}$ is a deterministic function that takes as input one element from G1, one element from G2, and output an element in group G3, which satisfies the following criteria. (1) Bi-linearity: For all $\mathrm{x} \in \mathrm{G1}, \mathrm{y} \in \mathrm{G2}, \mathrm{a}, \mathrm{b} \in \mathrm{Z_p}, \mathrm{e}(\mathrm{x^a}, \mathrm{y^b}) = \mathrm{e}(\mathrm{x}, \mathrm{y})^{\mathrm{ab}}$. (2) Non degeneracy: $e(g1, g2) \neq 1$ where $g1$ and $g2$ are generator of G1 and G2 respectively. (3) e must be computed efficiently.

Definition 2 (Discrete Logarithm Problem). Given two group elements g and h, find an integer a $\in \mathrm{Z_p}$ such that $h = g^a$ mod p whenever such integer exist.

3 The AB-OR Protocol

In this section we present our AB-OR protocol. The proposed approach consists of *6* algorithms. As said in previous section, the **Encrypt** and **Decrypt** algorithms were same as in [30], so it is omit in this section.

Setup: This algorithm run by CA (central authority) and it generates the public (MPK) and private (MSK) parameters. The MPK is available to all algorithms. It selects two multiplicative cyclic groups G and G1 with prime order p. g is the generator for G. It selects $y, \beta \in_R Z_p$. Bilinear map function e: $G \times G \rightarrow G_1$.

$$MPK = \{G, g, e, h = g^{\beta}, Y = e(g, g)^{y}\}$$
$$MSK = \{y, \beta\}$$

KeyGen (MSK,u): This algorithm run by CA to create the secret key (SK) for user/OR u. L is the list of attributes. Here we assume that L contains only one attribute i.e. *ID*. Therefore, in case of sender it is *S*, receiver it is *R* and for *OR-n* it is OR_n. H is the universal hash function H: {0, 1}* →G. CA selects $r, r_j \in_R Z_p$.

$$SK = (D = g^{(y+r)/\beta}, D_1 = g^r H(ID)^{r_j}, D_1' = g^{r_j})$$

Circuit_Construction (**MPK,M,OR**$_1$**,OR**$_2$**,...,OR**$_N$): This algorithm is run by sender. It will take message M as input. Here OR_1, OR_2, ..., OR_N is the list of intermediate ORs between sender and receiver. Sender make policy $W' = \{R\}$ and $CT' =$ **Encrypt**(MPK, M, W'). Now sender makes policy $W = \{OR_1\, or\, OR_2\, or \ldots or\, OR_N\, or\, R\}$ and $CT =$ **Encrypt**(MPK, CT', W). Sender sends the ciphertext CT to the OR_1.

Decrypt_ OR (SK,CT): The decryption algorithm takes *SK* of user and ciphertext *CT* as input. This algorithm will run at OR side as follows.

1. Onion Router i (OR$_i$) will get the *CT*.
2. If **Decrypt** *(SK,CT)* successful then send *CT* to all connected routers and users, otherwise simply discard the *CT*.
3. ORs cannot decrypt the *CT'* because they do not have attribute containing $id = R$.

Decrypt_ R (SK, CT): The decryption algorithm takes *SK* of user and ciphertext *CT* as input. This algorithm will run at receiver side as follows.
Get the message $M =$ **Decrypt**$(MPK, SK,$ **Decrypt**$(MPK, SK, CT))$ if SK satisfies CT and CT' otherwise "ϕ".

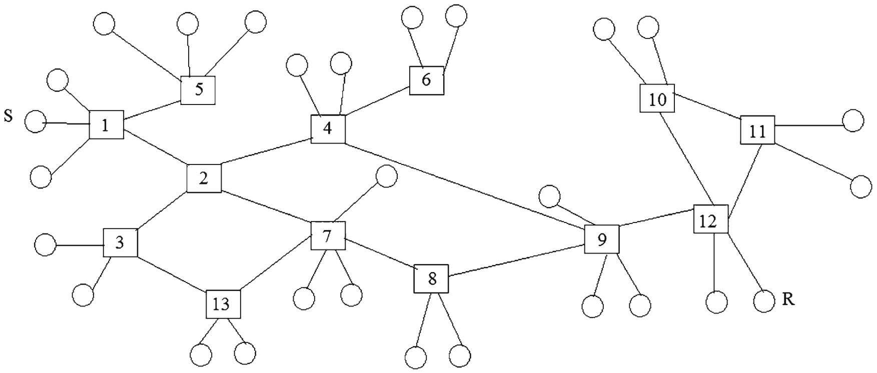


Fig. 2 Onion routing example

4 Analysis of Approach

In Fig. 2 the *round* represents the users and *rectangle* represents the ORs. *S* and *R* represent the sender and receiver respectively. Here we assume that each OR connected to more than 1 user to get better recipient anonymity. Assume that attacker compromised OR_7, so he only gets information that OR_2 was the previous hop in the path, but he cannot get the next hop in the path, this is a unique feature of our protocol which was not present in previous protocols.

Recipient Anonymity: In TOR protocol if attacker compromise OR_{12} then he knows the address or identity of the recipient which was not possible in our proposed protocol, any OR even does not know that whether the next hop in the path is an OR/recipient itself, so this give complete recipient anonymity. Let us take an example of battle field as discussed in Sect. 1. In our scheme base station send message to all the battlefield and only intended can able to decrypt, now enemy cannot imagine that which battlefield they have to fight/attack. This is one of the unique feature of AB-OR.

Forward secrecy: If the secret key or long term secret of OR_7 was compromised, even than attacker cannot predict about the path of past messages, so this will give the forward secrecy to our proposed scheme. If someone wants to use the concept of session keys [8] for forward secrecy than in AB-OR one can use the concept of fading function [21], in which the value of the attribute will be changed after certain time period.

Message consistency: It may look like the message is not changing at every hop in the path so this may give path information to an attacker, but for that attacker had to watch the entire network because (s)he do not know the next hop in the path. Therefore, for large network with thousands of network links this cannot be possible.

Network Overhead: At first sight it looks that network overhead will be more as we send messages to all except the receiving OR, but out of all only one OR can decrypt and the remaining will discard. In Fig. 2 we can say OR_7 receives message from OR_2 so OR_7 will send to OR_{13},OR_8 and three users. Only OR_8 in path so it can decrypt and the others will discard as they not able to decrypt. The CA will distribute the SKs to corresponding ORs initially.

Flooding: It may look that attacker compromise the honest node and send the continuous packet to other node to create a flooding. This will increase the traffic same as TOR protocol.

Computation Overhead:

1. *Sender side*: In TOR sender requires to do $n + 1$ time encryption if there are n ORs in the path. In proposed approach sender required to do 2 times encryption irrespective of number of *OR* in the path. Therefore, this will reduce the computational complexity from sender side. This is one of the unique feature of AB-OR.
2. *OR side*: They had to decryption *1 (one)* time as in *TOR*. There are *OR* which had to do the decryption although they are not in the path. In TOR, for example, asymmetric cryptography is used only to establish circuits, afterwards symmetric cryptography is used for forwarding traffic i.e. OR needs to run Diffie-Hellman key exchange protocol before communication can proceed. This overhead is not required in our scheme because circuit forms while the message is going on. By default, TOR requires 3 relays i.e. predecessor hop $\rightarrow$ current hop $\rightarrow$ successor hop. However, our scheme requires 2 relays because current hop does not require the address of successor hops as it forwards to all neighboring hope.
3. *Receiver side*: It requires to perform decryption two times. So only one more time than the *TOR*.

Path Discovery/Fault tolerance: Assume that *Sender* Ssends a message M to *Receiver* R using path S-OR_1-OR_2-OR_7-OR_8-OR_9-OR_{12}-R. What happen if OR_7 fails due to any reason? As per previous approaches there is no any other solution so the message is discarded after OR_2. In our proposed approach, we can make a policy in such a way so that failing of any OR will not affect the message. In the Fig. 2, if *Sender* not sure about OR_7 than he can make a policy $\mathrm{W} = \{\mathrm{OR}_1$ *or* OR_2 *or* OR_4 *or* OR_7 *or* OR_8 *or* OR_9 *or* OR_{12} *or* $\mathrm{R}\}$, where OR_k represent Onion Router k's ID and *or* represents ORing condition. Here we assume that *Sender* had sufficient knowledge of different paths to *the receiver*. This is one of the unique feature of AB-OR. It is important here that only *Sender* will decide the alternate path (and so the policy) not the intermediate nodes. If *sender* will aware that there are other senders exists in the network which sends messages to the same receiver than path discovery will become an important issue as the same path can lead to a congestion.

5 Conclusion and Future Work

In this paper we propose the new construction of onion routing called *AB-OR* and discuss the relative merits of it comparable to onion routing techniques till present. The disadvantage of proposed scheme was, it never changes the ciphertext between intermediate routers in the path so adversary may identify the path in network if he had access to all the connections in the network. In the future, we can enhance this protocol to change the ciphertext from each intermediate ORs. The security proof is given in selective secure that can be extending to full security model. If there are many ORs in path then size of final ciphertext will be increase. Therefore, one can use the concept of constant ciphertext length schemes to minimize network overhead and computation overhead at the decryptor side. The security proof and practical implementation will be given in the extended version of this paper.

References

1. Goldschlag D, Reedy M, Syverson P (1999) Onion routing for anonymous and private internet connections. Commun. ACM 42(2):39–41
2. Dingledine R, Mathewson N, Syverson P TOR: the onion router. Tor Project/EFF. http://www.torproject.org
3. Onion routing for anonymous communication. http://ntrg.cs.tcd.ie/undergrad/4ba2.05/group10/index.html
4. Dingledine R, Mathewson N, Syverson PF (2004) Tor: the second-generation onion router. In: USENIX security symposium, pp 303–320, 2004
5. Camenisch J, Lysyanskaya A (2005) A formal treatment of onion routing. In: Shoup V (ed) Proceedings of CRYPTO 2005. LNCS, vol 3621. Springer, Heidelberg, pp 169–187
6. Canetti R (2001) Universally composable security: a new paradigm for cryptographic protocols. In: Proceedings of the 42nd IEEE symposium on foundations of computer science 2001, pp 136–145
7. Kate A, Goldberg I (2010) Using sphinx to improve onion routing circuit construction. Financ Cryptogr Data Secur (LNCS) 6052:359–366
8. Catalano D, Fiore D, Gennaro R (2009) Certificate less onion routing. In: InCCS'09, pp 151–160
9. Kate A, Zaverucha GM, Goldberg I (2007) Pairing-based onion routing. In: 7th privacy enhancing technologies symposium (PETS 2007). Lecture notes in computer science, vol 4776. Springer, Heidelberg, pp 95–112
10. Kate A, Zaverucha GM, Goldberg I (2010) Pairing-based onion routing with improved forward secrecy. ACM Trans Inf Syst Secur 13(29):4
11. Klonowski M, Kutyłowski M, Lauks A (2008) Repelling detour attack against onions with re-encryption. Appl Cryptogr Netw Secur (LNCS) 5037:296–308
12. Borisov N, Klonowski M, Kutyłowski M, Lauks-Dutka A (2010) Attacking and repairing the improved modonions protocol. In: ICISC-09. LNCS, vol 5984. Springer, Berlin, pp 258–273
13. Gomułkiewicz M, Klonowski M, Kutyłowski M (2005) Onions based on universal re-encryption—anonymous communication immune against repetitive attack. Inf Secur Appl (LNCS) 3325:400–410
14. Ren J, Wu J (2010) Survey on anonymous communications in computer networks. Comput Commun 33(4):420–431

15. Danezis G, Diaz C (2008) A survey of anonymous communication channels. Technical report MSR-TR-2008-35. Microsoft Research, Cambridge, UK
16. Tang C, Goldberg I (2010) An improved algorithm for Tor circuit scheduling. Technical report CACR 2010–06, University of Waterloo
17. Johnson N, McLaughlin S, Thompson J (2010) Path tracing in TOR networks. In: 18th European signal processing conference (EUSIPCO-2010), pp 1856–1860. ISSN 2076–1465
18. Feigenbaum J, Johnson A, Syverson P (2007) A model of onion routing with provable anonymity. In: Proceedings of the 11th financial cryptography and data security conference (FC 2007)
19. Camenisch J, Neven G (2010) Saving on-line privacy. In: IFIP advances in information and communication technology, vol 320. Springer, Boston, pp 34–47
20. Kaviya K (2009) Network security implementation by onion routing. In: Proceedings of the 2009 international conference on information and multimedia technology (ICIMT '09). IEEE Computer Society, Washington DC, pp 339–342
21. Panchenko A, Pimenidis L, Renner J (2008) Performance analysis of anonymous communication channels provided by Tor. In: Proceedings of the third international conference on availability, reliability and security (ARES 2008), Barcelona. IEEE Computer Society Press, Washington DC, pp 221–228
22. Snader R, Borisov N (2008) A tune-up for Tor: improving security and performance in the Tor network. In: Proceedings of the network and distributed security symposium—NDSS '08
23. Panchenko A, Renner J (2009) Path selection metrics for performance-improved onion routing. In: Proceedings of the 9th IEEE/IPSJ symposium on applications and the internet (IEEE SAINT 2009), Seattle, July 2009. IEEE Computer Society Press, Washington DC
24. Catalano D, Di Raimondo M, Fiore D, Gennaro R, Puglisi O (2011) Fully non-interactive onion routing with forward-secrecy. In: Proceedings of the 9th international conference on applied cryptography and network security (ACNS'11). LNCS, vol 6715. Springer, Berlin, pp 255-273
25. Backes M, Goldberg I, Kate A, Mohammadi E (2011) Provably secure and practical onion routing. Cryptology ePrint Archive, Report 2011/308
26. Egners A, Gatzen D, Panchenko A, Meyer U (2012) Introducing SOR: SSH-based onion routing. In: Proceedings of the eighth international IEEE symposium on frontiers of information systems and network applications (FINA-2012), part of the 26th IEEE international conference on advanced information networking and applications (IEEE AINA 2012). IEEE Computer Society Press, Washington DC
27. Patil NM, Lingam C (2012) Anonymus connections and onion routing. Int J Adv Res Comput Sci Softw Eng 2(2). ISSN: 2277–128X
28. Sahai A, Waters B (2005) Fuzzy identity-based encryption. In: Proceeding of the EUROCRYPT. LNCS, vol 3494. Springer, Berlin, pp 457–473
29. Goyal V, Pandey O, Sahai A et al (2006) Attribute based encryption for fine-grained access control of encrypted data. In: Proceedings of the 13th ACM conference on computer and communications security. ACM, New York, pp 89–98
30. Bethencourt J, Sahai A, Waters B (2007) Ciphertext-policy attribute-based encryption. In: Proceedings of the 2007 IEEE symposium on security and privacy (S&P 2007). IEEE, Piscataway, pp 321–334
31. Cheung L, Newport C (2007) Provably secure ciphertext policy ABE. In: Proceedings of the 14th ACM conference on computer and communications security. ACM, New York, pp 456–465
32. Zhibin Z, Dijiang H (2012) On efficient ciphertext-policy attribute based encryption and broadcast encryption: extended abstract. In: Proceedings of the 17th ACM conference on computer and communications security (CCS '10). ACM, New York, pp 753–755
33. Emura K, Miyaji A, Nomura A, Omote K, Soshi M (2009) A ciphertext-policy attribute-based encryption scheme with constant ciphertext length. In: Bao F, Li H, Wang G (eds) ISPEC 2009. LNCS, vol 5451. Springer, Heidelberg, pp 13–23
34. Paterson K, Quaglia E (2010) Time-specific encryption. In: Garay J (ed) Proceedings of seventh conference on security and cryptography for Networks

Pasic: A Novel Approach for Page-Wise Web Application Security

Angamuthu Maheswaran and Rajaram Kanchana

Abstract Secured access to Web contents and the interaction with Web application are becoming one of the most important issues in the context of Internet. HTTP protocol which uses plain text transmission is employed for data communication over Internet. Secure Socket Layer (SSL) certificates over HTTP evolve into HTTPS protocol which is one of most used solutions that provide security. However the same certificate has been used for all the pages irrespective of sensitivity of the data. Moreover, data with different security requirements have been secured using the same algorithm which could either reduce the performance of the Web application or do not provide the appropriate security according to the nature of each data item. In order to compensate the degradation in the quality of service, it is proposed to use appropriate encryption and integrity algorithms for each page, based on the sensitivity of information and security requirements for the data. A gradation of security levels namely high, medium, and low has been proposed. A combination of different algorithms are considered to provide confidentiality and integrity for each level of security. The proposed approach is experimented with a prototype in healthcare domain.

Keywords Secure scoket layer · Healthcare information management sysytem · Web application security · Confidentiality · Integrity

A. Maheswaran (✉) · R. Kanchana
Department of Computer Science and Engineering, SSN College of Engineering,
Anna University of Technology, Chennai 603110, Tamil Nadu, India
e-mail: mahe.karthick@gmail.com

R. Kanchana
e-mail: rkanch@ssn.edu.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_43,

1 Introduction

Web application usage has increased more due to internet and lot of information has been shared between client and Web application. Web applications need to provide certain levels of security so that the user develops trust upon the application while providing input data. Plain text communication of data across the Web is fine for Web sites holding less sensitive information such as news and sports information. However, unsecured communication of more sensitive information such as Social Security Number (SSN), credit card details, and cryptographic key becomes unacceptable, since third parties would use tools like network-sniffer to eavesdrop the transmitted data. In internet, Web browser and Web server communicate using HTTP protocol. HTTP is capable of transmitting huge volume of data as plain text, in short time. However it is vulnerable for attacks and does not guarantee data integrity. SSL [1] protocol evolved as a solution for addressing the shortcomings in the existing HTTP mode of communication. However, Security in existing Web applications is provided using a single SSL certificate, irrespective of sensitivity of data in the application [2].

Normally different data items communicated to the Web server are functionally related and obtained from a single page. Hence, it is assumed that each communication involves data belonging to a single page. All the pages in a Web application do not have data with same sensitivity. Applying same security algorithm for all the Web pages may increase the application execution time, if only few of the Web pages contain high sensitive data and most complicated security algorithm is applied to all the Web pages. When less complicated security algorithm is applied to all the Web pages, the high sensitive data may not be communicated with correct security mechanism. To address this problem, a novel approach named as PASIC (Page-Wise Security for Integrity and Confidentiality) is proposed for providing page-wise security based on the sensitivity or defined security of the data communicated from a Web page. Three levels of security have been proposed namely low, medium, and high. Each Web page is classified under one of these security levels based on the security policies in the business application. Further, the available security algorithms for confidentiality and integrity are assigned to each of these levels based on the robustness and complexity of the algorithms.

2 Related Work

In general, the performance of SSL/TLS servers decreases due to increased user requests or Denial of Services (DoS) attacks. Luo et al. [3] investigated the performance and security of three algorithms in SSL/TLS handshake protocol. However, some algorithms have very specific application domains and the authors have summarized the disadvantages of each algorithm. Huawe et al. [4] have proposed two schemes such as a modified handshaking protocol and a Key generation protocol in order to strengthen SSL security. However, the authors assume that the wireless channel is secure and the modified approach does not provide security in wireless channel. With server initiated authentication protocols, server authentication may not be performed and hence, attacks such as phishing, pharming, and Man-In-The-Middle (MITM) can easily be deployed. In order to overcome this problem, Masaru [5] proposed a scheme for user initiated server authentication and two schemes for protecting against Cross-Site Scripting (XSS) and Cross-Site Reference Forgery (XSRF). These scheme have been integrated into the HTTP based system and the browser has been extended. This integration requires moderate work and the performance of approach is yet to be assessed with benchmark and attack databases.

Though existing approaches have attempted to implement SSL and improve the performance of SSL protocol, these approaches use the same algorithm for communication of all the data items in a Web application irrespective of their sensitivity. In contrast, the proposed approach, PASIC, uses different levels of security and different kinds of algorithms for each security level, for the set of data items communicated from a Web page, based on their sensitivity and defined security.

3 System Architecture

A three layer architecture comprising of application layer, server middle ware, and storage layer has been adopted by the proposed approach PASIC. While providing the required services to the user, the server provides several controllers that are useful in secured communication. The architecture for PASIC, shown in Fig. 1, makes use of *PASICSecurityManager* to achieve secured communication by integrating the security features in the controller. The sensitivity of data items or a Web page is specified in Security Level Configuration database, *SecurityLevelConfig*. Three levels of security have been proposed namely low, medium, and high. For each security

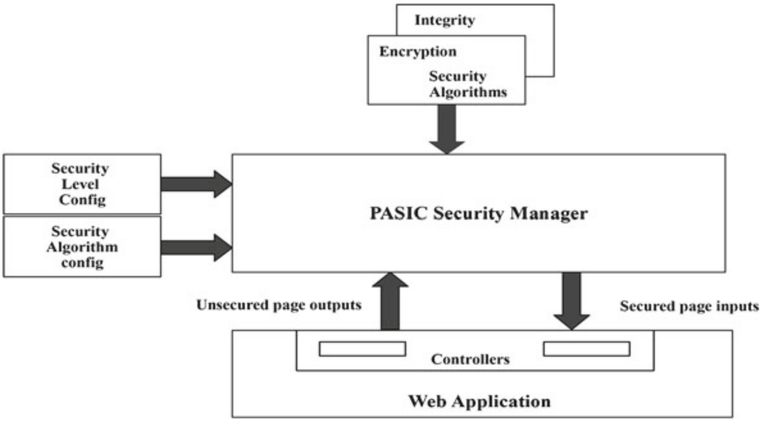


Fig. 1 System architecture

level, Encryption/Decryption and Integrity algorithms are configured in Security Algorithm Configuration database, SecurityAlgorithmConfig. The security library consists of a list of security algorithms. The security library is built with various security algorithms for encryption, such as EDH (Ephemeral Diffe-Hellman), DSS (Digital Signature Standard), RSA (Rivest Shamir Adleman), RC4 (Ron's Code 4), DES (Data Encryption Standard), and AES (Advanced Encryption Standard). It also contains algorithms for integrity such as SHA1 (Secure Hashing Algorithm 1) and MD5 (Message Digest 5).

The security manager integrates security level configuration, security algorithm configuration, and the security library. It also provides a secure environment for data communication. While data is sent from a Web page to the Web server, the sensitivity level of data is obtained from the *SecurityLevelConfig* database and the corresponding algorithm configured in the *SecurityAlgorithmConfig* database is chosen. The data is encrypted using the chosen security algorithms available in the security library before the data is communicated. When the data is received in the destination page, it is decrypted using the same algorithm.

4 Prototype

A prototype of Healthcare Information Management System (HIMS), for paper less work in a hospital has been built to validate the proposed PASIC approach. The prototype is developed with seven roles of the user namely: Administrator, Consultant, Scan Centre, Main Pharmacy, In Patient Pharmacy, Out Patient Pharmacy, and Nurse Station. Various user roles and operations supported for each role are depicted in Fig. 2. Administrator configures the security levels of each page and different

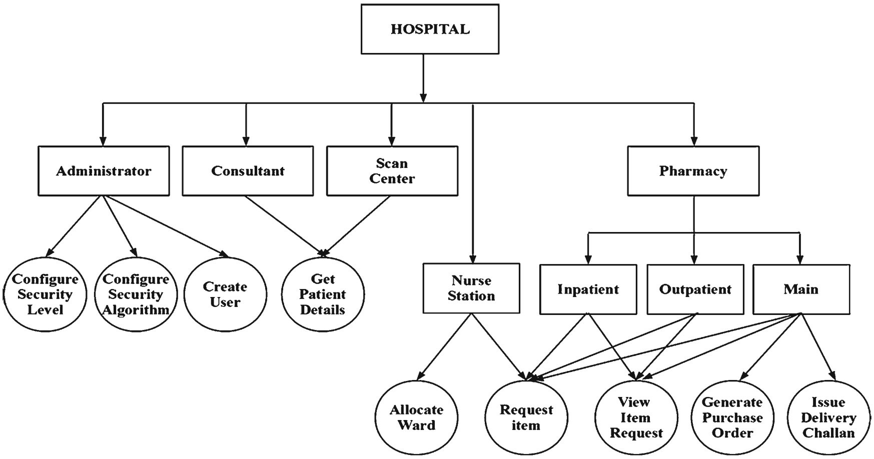


Fig. 2 HIMS prototype

combination of algorithms for each security level, while also creating user name and password for new employees. The consultant obtains the details of symptoms from patients and suggests diagnosis along with scan requests, if necessary. The scan center employee views the scan requests and acquires the scan for patient. The Main pharmacy places orders for medicines and issues a delivery Challan upon receiving the medicines. The users of Main Pharmacy, Inpatient pharmacy, and Outpatient pharmacy can request items among themselves and view the item requests. The wards are allocated to patients by a nurse in the Nurse Station. The nurse can also request items from pharmacies and view the details of wards allocated to patients.

5 Implementation

The HIMS prototype has been implemented on Ruby On Rails (ROR) [6] using object relational programming language, Webrick server, SQLite3 database, and OPENSSL library with RadRails IDE. It runs on 2.4 GHz Intel CORE i3 processor with 3 GB RAM. Ruby on Rails uses MVC architecture [7] and is capable of implementing models, views, and controllers. The controller is used to control the flow of execution among the Web pages and between a Web page and the databases. Models are used to store the data and views are used to provide user interface. OPENSSL library has a list of cryptography algorithms such as AES-256, DES-CBC, and DES3 used for encryption and hashing algorithms such as SHA1 and MD5 [8, 9] used for integrity. The HIMS Web application needs to be accessed in various departments of the hospital. Various users of the application are connected through Intranet along with the Web server in the hospital. The encryption controller is used to encrypt the data using the algorithms available in OPENSSL library. While the data is sent from a Web page, the control is delegated to the encryption controller which would perform the required action. The security level of the page is determined from the security level configuration database. Based on the security level, corresponding encryption and integrity algorithms are chosen from the security algorithm configuration database. The data obtained from the Web page are encrypted and hash values are calculated. The encrypted data and hash value are sent to the Web server. The control is then delegates to the decryption controller to decrypt the data using the algorithms chosen from the configuration databases. The hash values are calculated from the decrypted data and if the received hash value matches with the calculated hash value, the data is processed by the Web server. Whenever hash values do not match, it means that there exists data loss and hence, an error message would be displayed. After processing the data, the encryption controller encrypts and sends them to target Web page. Before rendering into the target Web page, the decryption controller decrypts the data. Whenever the data needs to be communicated from a source Web page to a target Web page without any processing, the data is encrypted and decrypted once. Only when the there is a necessity to process the data, an additional encryption and decryption are required.

Table 1 Prototype security level configuration

Web page number	Web page	Security level
1	SignIn	High
2	Consultant diagnosis	High
3	Scan center	Medium
4	Main pharmacy	Medium
5	Main pharmacy purchase order	Low
6	Main pharmacy delivery challan	Medium
7	Main pharmacy item requistion	High
8	Main pharmacy view request	Low
9	IP pharmacy item request	Low
10	OP main pharmacy item request	Low
11	Nurse form	High
12	Nurse item request	Medium

Table 2 Security algorithm configuration 1

Security level	Encryption/Decryption	Integrity
Low	DES3	SHA1
Medium	RSA	MD5
High	Modified RSA	MD5

5.1 Configuration of Security Levels and Algorithms

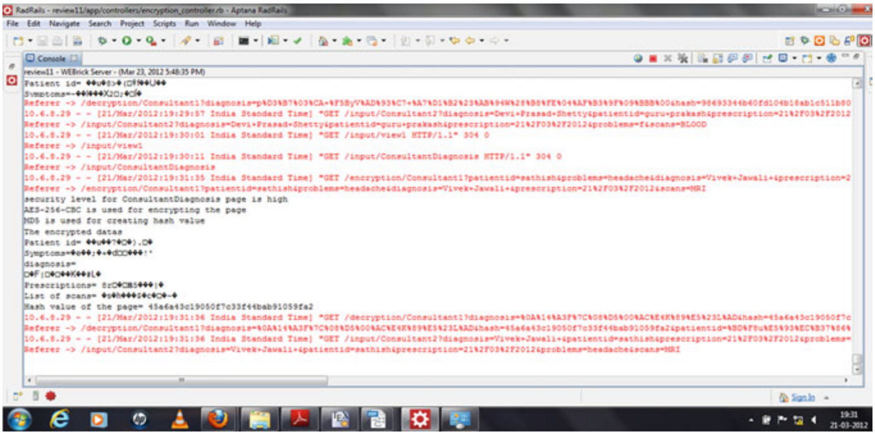
In order to secure the content, admin user configures the security level of each page, based on sensitivity of data contained in that Web page. In addition, for each security level the encryption and integrity algorithms for encryption and hashing are specified. A sample configuration of security levels and algorithms are shown in Tables 1 and 2. The Table 1 shows the security levels of HIMS Web pages according to the sensitivity of information presented in Web page.

5.2 Test Scenario

In order to demonstrate the proposed PASIC approach, the following scenario has been considered in the HIMS system.

- *Consultant diagnosis* page where the communicated data is encrypted and decrypted once, since, the data flows from source page to Web server and gets stored in the database.

While accessing the HIMS prototype, after successfully login, the user selects an appropriate role. The Web page used by the consultant has details about symptoms, diagnosis for the disease, and scan requests for future treatment. The data received


RadRails - review11/app/controllers/encryption_controller.rb - Aptana RadRails
File Edit Navigate Search Project Scripts Run Window Help
Console
Sign In
19:31
21-03-2012

```
review11 - WEBrick Server - (Mar 23, 2012 5:48:35 PM)
Patient id= ��u��>�(□f��U��
Symptoms=-�����X2□;�□�
Referer -> /decryption/Consultant1?diagnosis=p%D3%B7%03%CA+%F5By V%AD%93%C7+%A7%D1%B2%23%AB%94W%28%B8%FE%04%AF%B3%9F%09%BBB%00&hash=98693344b60fd104b18ab1c511b80
10.6.8.29 - - [21/Mar/2012:19:29:57 India Standard Time] "GET /input/Consultant2?diagnosis=Devi+Prasad+Shetty&patientid=guru+prakash&prescription=21%2F03%2F2012
Referer -> /input/Consultant2?diagnosis=Devi+Prasad+Shetty&patientid=guru+prakash&prescription=21%2F03%2F2012&problems=f&scans=BLOOD
10.6.8.29 - - [21/Mar/2012:19:30:01 India Standard Time] "GET /input/view1 HTTP/1.1" 304 0
Referer -> /input/view1
10.6.8.29 - - [21/Mar/2012:19:30:11 India Standard Time] "GET /input/ConsultantDiagnosis HTTP/1.1" 304 0
Referer -> /input/ConsultantDiagnosis
10.6.8.29 - - [21/Mar/2012:19:31:35 India Standard Time] "GET /encryption/Consultant1?patientid=sathish&problems=headache&diagnosis=Vivek+Jawali+&prescription=2
Referer -> /encryption/Consultant1?patientid=sathish&problems=headache&diagnosis=Vivek+Jawali+&prescription=21%2F03%2F2012&scans=MRI
security level for ConsultantDiagnosis page is high
AES-256-CBC is used for encrypting the page
MD5 is used for creating hash value
The encrypted datas
Patient id= ��u��7�□�).□�
Symptoms=�e��;�+�d□□���!'
diagnosis=
□�F|□�□��K��#L�
Prescriptions= 8r□�□B5���|�
List of scans= �s�h���$�c�□�-�
Hash value of the page= 45a6a43c19050f7c33f44bab91059fa2
10.6.8.29 - - [21/Mar/2012:19:31:36 India Standard Time] "GET /decryption/Consultant1?diagnosis=%0A%14%A3F%7C%08%D5%00%AC%E4K%89%E5%23L%AD&hash=45a6a43c19050f7c
Referer -> /decryption/Consultant1?diagnosis=%0A%14%A3F%7C%08%D5%00%AC%E4K%89%E5%23L%AD&hash=45a6a43c19050f7c33f44bab91059fa2&patientid=%BD%F8u%E5%93%EC%B37%86%
10.6.8.29 - - [21/Mar/2012:19:31:36 India Standard Time] "GET /input/Consultant2?diagnosis=Vivek+Jawali+&patientid=sathish&prescription=21%2F03%2F2012&problems=
Referer -> /input/Consultant2?diagnosis=Vivek+Jawali+&patientid=sathish&prescription=21%2F03%2F2012&problems=headache&scans=MRI
```

Fig. 3 Encryption and hashing of consultant diagnosis page

from this Web page is classified to be in high security level. After consultant inputs the details, the encryption controller is invoking to encrypt the data. The security level of Web page, security and integrity algorithms, encrypted data, and the hash value details are shown in Fig. 3. In this scenario, the diagnosis details flow from Consultant diagnosis page to Web server and gets stored in the diagnosis database. Hence, the communicated data is encrypted and decrypted once. After patient details entry, diagnosis database, patient database, scan request database all are updated.

5.3 Performance Evaluation

In order to assess the performance of PASIC approach for providing page-wise security, it is tested on HIMS prototype. The execution time consumed by the HIMS prototype, where each page is configured with defined security level is measured. Three configurations for security levels have been considered wherein different algorithms are assigned for different security levels. The performance of HIMS prototype is also measured by configuring every page with the same security level (high, medium, or low). The performance of HIMS with PASIC approach applying security algorithms DES-3, SHA1 for low level, RSA, MD5 for medium level, Modified-RSA, MD5 for high level is shown in Table 4. The first three columns of the table shows the performance of HIMS applying DES-3, SHA1 for all pages, RSA, MD5 for all pages, and Modified-RSA, MD5 for all pages respectively. Each measurement has been computed as an average of five test runs. The comparison of these four ways of providing security to HIMS prototype is shown in graph depicted in Fig. 4.

Similarly, the performance of HIMS has been measured for one more security level configuration 2 shown in Table 3 and are plotted in a graph shown in Fig. 5.

Table 3 Security algorithm configuration 2

Security level	Encryption/Decryption	Integrity
Low	DES3	SHA1
Medium	IDEA	MD5
High	RSA	MD5

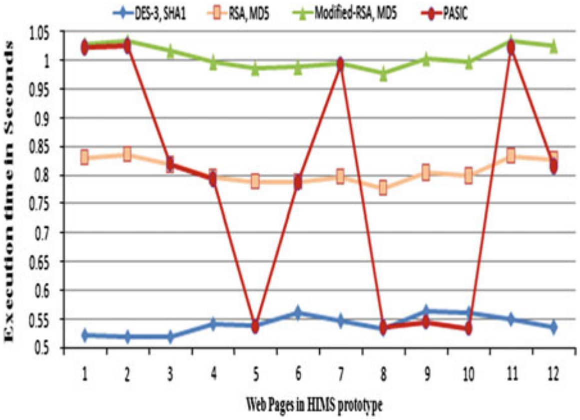


Fig. 4 Performance comparison of PASIC—configuration 1

The total execution time of HIMS for three different configurations is tabulated in Table 5.

It is observed from the graph that the proposed page-wise security approach executes faster than providing same higher level security for all the pages of HIMS prototype. It is observed that the approach of applying same security algorithm for all the pages does not provide the defined security. Whereas the proposed PASIC approach provides appropriate security defined for each of the Web pages while keeping the performance intact. The total execution time of HIMS required to adapt the proposed PASIC approach is optimal. The performance of HIMS prototype without providing any security has been measured as 2.96 s. Table 5 also shows the performance comparison of HIMS by providing same security for all pages and by providing page-wise security against without providing any security mechanism. For example, while applying DES-3, SHA1 algorithms for all the Web pages the total execution time is found to be 6.48 s which implies there is an increase of 118.91 % in execution time for providing security. Alternatively, with configuration 1, the proposed PASIC approach incurs 9.42 s. Thus it is evident that the proposed PASIC approach provides defined security for each Web page based on the business requirements with optimal performance.

Table 4 Performance comparison of PASIC—configuration 1

Web page number	DES-3, SHAI	RSA, MD5	Modified-RSA, MD5	PASIC with configuration 1	
1	0.522675	0.829696	1.029696	High	1.021698
2	0.520182	0.835208	1.035208	High	1.025098
3	0.518931	0.817699	1.017699	Medium	0.818291
4	0.540933	0.796885	0.996885	Medium	0.792487
5	0.539187	0.787694	0.987694	Low	0.536683
6	0.559434	0.788206	0.988206	Medium	0.787137
7	0.545932	0.795674	0.995694	High	0.992675
8	0.532182	0.776696	0.976696	Low	0.534681
9	0.564684	0.803692	1.003698	Low	0.543687
10	0.560683	0.797698	0.997698	Low	0.532682
11	0.550683	0.833694	1.033694	High	1.021694
12	0.534482	0.825499	1.025499	Medium	0.814846
Execution time in seconds	6.489988	9.688361	12.08837		9.421675

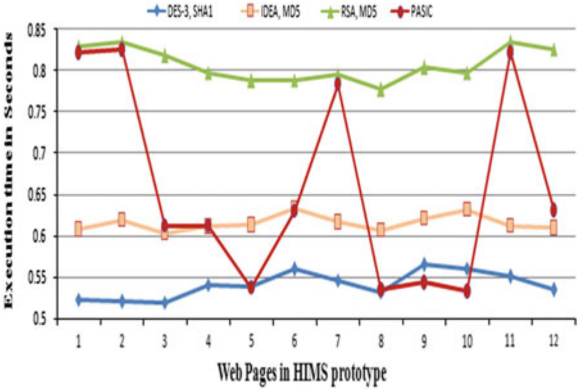


Fig. 5 Performance comparison of PASIC—configuration 2

6 Conclusion

A page-wise security approach is proposed for providing security to Web applications. The proposed approach provides defined security for every set of data items that is communicated between Web pages. Rather than applying the same security algorithm for any data to be communicated without considering its sensitivity, the proposed PASIC approach enables configuring different security levels and algorithms based on security requirements of data. A secured Web application for Healthcare Information Management System prototype is developed using the proposed PASIC approach. Its performance is measured and compared with existing SSL approach

Table 5 Execution time of HIMS with different PASIC configuration

Same security for all pages									PASIC		
Low level security			Medium level security			High level security			With page wise Security		
Algorithm	Execution time	% increase	Algorithm	Execution time	% increase	Algorithm	Execution time	% increase	configuration	Execution time	% increase
DES-3, SHAI	6.48	118.91	RSA, MD5	9.68	227.02	Modified RSA, MDS	12.08	308.18	Configuration 1	9.42	218.24
DES-3, SHAI	6.48	118.91	IDEA, MD5	7.38	149.32	RSA, MDS	9.68	227.02	Configuration 2	7.88	166.21
DES-3, SHAI	6.48	118.91	DES-ECB, MD5	6.51	119.93	AES-256 CBC, MDS	6.66	125	Configuration 3	6.58	122.29

of providing single security algorithm for every Web page. The PASIC approach provides optimal performance of the Web application in comparison to the existing SSL based approach in addition to defined adequate security for each data. The scalability of the PASIC approach needs to be assessed by implementing it in a more complex healthcare application involving several functionalities. The ongoing work considers development of more complex security algorithms in medium and high security levels to assess the performance of PASIC approach.

References

1. Stallings W (2002) Cryptography and network security-principles and practice, 3rd edn. Prentice Hall, Engle-wood Cliffs
2. Ben G, Whitney H, Andre H, Murali J, Prasad DV, Ravi T, David W (2002) Professional Web Services Security. Shroff Publishers and Distributors, Professional
3. Luo Q, Lin Y (2009) Analysis and comparison of several algorithms in SSL/TLS handshake protocol. In: IEEE international conference on information technology and computer science, pp 613–617
4. Huawei Z, Ruixia L (2009) A scheme to improve security of SSL. In: Proceedings of the 2009 pacific-asia conference on circuits communications and system, pp 401–404
5. Masaru T (2009) An HTTP extension for secure transfer of confidential data. In: IEEE international conference on networking architecture and storage, pp 101–108
6. Fisher T (2008) Ruby on rails bible. Wiley Publishing Inc, New York
7. Model-view-controller architecture. http://www.jcorporate.com/expresso/doc/edg/edgWhatIsMVC.htmls
8. Openssl security implementation. http://www.ruby-forum.com/topic
9. Openssl algorithms. http://stackoverow.com/questions/2043557/des3-decryption-in-ruby-on-rails

Cryptanalysis of Lo et al.'s Password Based Authentication Scheme

Nishant Doshi and Bhavesh Patel

Abstract A key exchange protocol allows more than two parties to communicate over the insecure channel to establish common shared secret key called session key. Due to the significance of this notion to establish secure communication among parties, in literature there have been numerous approach have been proposed and analyzed based on their merits and de-merits. Recently, Lo et al. proposed a 3-party Password based Authenticated Key Exchange protocol in which two or more users equipped with pre-shared secrets to the server and can able to generate the session key with the help of the server. They claimed that their approach is resist against any known attacks. However, we observe that their protocol is not secure against against off-line password guessing attack, long term secret compromise attack as well as compromise of previous session can lead to compromise all involving users for future communication. Therefore, in this this paper first we have analyzed these attacks and suggest the improve scheme that overcomes these attacks.

Keywords Attack · Cryptanalysis · Offline password · Key exchange · Authendication

1 Introduction

In this digital communication world, every user is connected through the internet called an insecure channel in the sense that any intermediate user can able to track your data which are sent. In a variety of applications, there is a need to setup a shared

N. Doshi (✉)
National Institute of Technology, Surat, India
e-mail: doshinikki2004@gmail.com

B. Patel
S N P I T & R C, Vidyabharti Campus, Umrakh, India
e-mail: bhavesh.patel17@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_44,

common secret key among a particular set of users for a particular time i.e. session key. The session key is useful to exchange the messages between two or more users securely over insecure channel.

For example, defense minister wants to communicate prime minister and home minister to discuss security issues related to the country. They can generate a common shared secret key and continue over an insecure channel. Like a this there are many applications in a real life that uses the authentication schemes to share a common shared secret key.

To address this issue, in literature, a password based authentication key exchange (PAKE) scheme is proposed in which two or more user authenticates each other, generate a session key over a secure channel with the help of a central trustable server.

In [1] authors firstly address this issue and proposed two party password based authentication scheme (2PAKE). Thereafter many solutions proposed to improve upon a basic scheme in literature [2–24]. However, in 2PAKE protocols, if the number of users increases than the number of shared password also increases which can lead to bottleneck problem in real time scenario.

Therefore, in [8] authors firstly proposes the 3PAKE protocol to reduce the burden on each individual party. In [25] the authors prove that the scheme of [8] is vulnerable to on-line password guessing attack. Same as this in [26] authors prove the attack of offline password guessing on [25]. As the scheme of [26] using Public Key Cryptosystem (PKC), it requires high computation power in real time scenario. Therefore, in [27] authors proposed 3PAKE scheme without PKI while in [28] authors have used the Weil pairing. In [29], the authors prove that the scheme of [28] is vulnerable to Main-in-Middle attack. Afterwards there have been many schemes [30–45] proposed in the literature with their merits and de-merits.

Recently Lo et al. [46] proposed the scheme for 3PAKE and claim that their scheme is secured against any known attacks. However, as we observed that Lo et al.'s scheme is vulnerable to some attacks including offline password guessing attack. The details cryptanalysis of this attack is given in Sect. 2. Thereafter we have proposed the new improved scheme to overcome these attacks.

1.1 Organization of the Paper

In Sect. 2, we have given the preliminary and thorough analysis of the attacks on Lo et al.'s scheme. In Sect. 3, we have given the proposed scheme. The conclusion and scope for future work is given in Sect. 4. The references are given at the end.

2 Cryptanalysis of Lo et al.'s [46] Scheme

In this section we have given the preliminaries that we will use throughout the paper as well as the attacks on the Lo et al.'s scheme.

2.1 Preliminaries

Symbol	Meaning
A,B	Two communicating parties
S	The trusted server
IDA	The identiy of party A
P_A, P_{A1}	The passwords shared between party A and server S. Same for B
$E_k(M)$	The message M is encrypted using symmetric key k
p	The large prime number and order of group G
g	The generator for group G
$H()$	The universal one way collusion resistance hash function
Z_p	A set of elements of range 0…p-1
$PRF_k()$	A pseudo-random function with three parameters and indexed by key k

2.2 Cryptanalysis

First we will take a look at the working of the Lo et al.'s scheme. Here A and B wants to communicate with S and generate the shared secret key. The working of scheme is as follows. A generates $r_A \in_R Z_p$ and computes $N_A = g^{r_A}\ mod\ p$. Then sends the message $\{ID_A, ID_B, E_{p_A}(N_A)\}$ to S. B generates $r_B \in_R Z_p$ and computes $N_B = g^{r_B}\ mod\ p$. Then sends the message $\{ID_A, ID_B, E_{p_B}(N_B)\}$ to S.

After receiving messages from A and B, S gets N_A and N_B using shared secret key P_A and P_B respectively on $E_{p_A}(N_A)$ and $E_{p_B}(N_B)$. S generates $r_S \in_R Z_p$ and compute $N_A^{r_S} = g^{r_S r_A}\ mod\ p$ and $N_B^{r_S} = g^{r_S r_B}\ mod\ p$. Then S computes $K_{AS} = N_A^{p_{A1}} = g^{p_{A1} r_A}\ mod\ p$ and $K_{BS} = N_B^{p_{B1}} = g^{p_{B1} r_B}\ mod\ p$. S sends the $X_A = g^{r_B r_S} f_{K_{AS}}(ID_A, ID_B, N_A)$ to A as well as $X_B = g^{r_A r_S} f_{K_{BS}}(ID_A, ID_B, N_B)$ to B.

After receiving messages from S, A computes $K_{AS} = N_A^{p_{A1}} = g^{p_{A1} r_A}\ mod\ p$, current session key $K = \left(\frac{X_A}{fk_{AS}(ID_A, ID_B, N_A)}\right)^{r_A}\ mod\ p = g^{r_A r_B r_S}\ mod\ p$. A sends the verification message $\alpha_A = H(ID_A, ID_B, K)$ to B. Same as A, the B computes $K_{BS} = N_B^{p_{B1}} = g^{p_{B1} r_B}\ mod\ p$, current session key $K = \left(\frac{X_B}{fk_{BS}(ID_A, ID_B, N_B)}\right)^{r_B}$ $mod\ p = g^{r_A r_B r_S}\ mod\ p$. B sends the verification message $\alpha_B = H(ID_A, ID_B, H(K))$ to A. After verification at both sides, the current session key K is accepted for future communication between A and B.

2.2.1 Offline Password Guessing Attack

In the real time scenario, a user selects the password that is easily remembered as they are related to the user by some means say Birthdate, SSN number, college ID number etc. From this notion the attacker can guess the password and launch the attack after communication called offline guessing attack.

Table 1 Offline password guessing attack

Offline password guessing attack $(\boldsymbol{X_A}, \boldsymbol{X_B}, \boldsymbol{D_A}, \boldsymbol{D_B})$

```
For $i := 0$ to $|D_A|$
For $j := 0$ to $|D_B|$
$P^*_{A1} \leftarrow D_i; P^*_{B1} \leftarrow D_j$
compute $K^*_{AS} = N_A^{p*_{A1}} \ mod \ p \ and \ K^*_{BS} = N_B^{p*_{B1}} \ mod \ p$
compute $X^*_A = \frac{X_A}{fK^*_{AS}(ID_A, ID_B, N_A)} \ and \ X^*_B = \frac{X_B}{fK^*_{AS}(ID_A, ID_B, N_B)}$
if $(X^*_A) r_B^{-1} \ mod \ p = (X^*_B)^{r_A^{-1}} \ mod \ p$
then return $P^*_{A1}, P^*_{B1}$
```

Let us assume that attacker C has control over the communication channel. C first guess the P_A and try to validate based on decryption of N_A. Let D be the set of candidate password. If we assume that there are n communications taken place than C can guess the password with probability $1 - \left(\frac{1}{2}\right)^n$ which is negligible with larger n. After getting P_A and P_B, C can get the P_{A1} and P_{B1} as follows. Here C will start a new session with S on behalf of A and B so C knows r_A, r_B, N_A, N_B. Here D_A and D_B the password guessing space for A nad B respectively Table 1.

2.2.2 Impersonate Attack

Here we assume that attacker C have the following values from the past session.

- $r_A, E_{P_A}(N_A), K_{AS} = N_A^{P_{A1}} \ mod \ p$

It is clearly seen that with values C can not get the value of P_{A1} because of discrete logarithm problem. Now we can show that without using the technique for offline password guessing attack, C can impersonate as A to B with any future session as follow.

- C will use the same value i.e. r_A, $E_{p_A}(N_A)$ for new session with S.
- B generates the new value say r'_B and so on for session.
- In the round-3, C gets $X_A = g^{r'_B r_S} f_{k_{AS}}(ID_A, ID_B, N_A)$ and computes the session key $K = \left(\frac{X_A}{fk_{AS}(ID_A, ID_B, N_A)}\right)^{r_A} \ mod \ p$ i.e same as A can compute in previous session using K_{AS} and r_A.

3 The Proposed Scheme

Here user A generates P_A, $P_{A1} \in_R Z_p$ and submit to S using secure channel. Same for user B.

- Round-1
  - A generates $r_A \in_R Z_p$ and computes $N_A = g^{r_A} \; mod \;\; p$. Then sends the message $\{ID_A, ID_B, E_{p_A}(N_A)\}$ to S.
  - B generates $r_B \in_R Z_p$ and computes $N_B = g^{r_B} \; mod \;\; p$. Then sends the message $\{ID_A, ID_B, E_{p_A}(N_B)\}$ to S.
- Round-2
  - After receiving messages from A and B, S gets N_A and N_B using shared secret key P_A and P_B respectively on $E_{p_A}(N_A)$ and $E_{p_B}(N_B)$.
  - S generates $r_s, R_S \in_R Z_p$ and compute and $N_A^{r_s} = g^{r_S r_A} \; mod \;\; p$ and $N_B^{r_s} = g^{r_S r_B} \; mod \;\; p$.
  - Then S computes $K_{AS} = N_A^{p_{A1}} = g^{p_{A1} r_A} \; mod \;\; p$ and $K_{BS} = N_B^{p_{B1}} = g^{p_{B1} r_B} \; mod \;\; p$.
  - S sends the $X_A = g^{r_B r_S} f_{k_{AS}}(ID_A, ID_B, N_A)$, $X'_A = R_s^{p_{A1}} \; mod \;\; p$ to A as well as to $X_B = g^{r_A r_S} f_{k_{BS}}(ID_A, ID_B, N_A)$, $X'_B = R_s^{p_{B1}} \; mod \;\; p$ to B.
- Round-3
  - After receiving messages from S, A computes, $K_{AS} = N_A^{P_{A1}} = g^{p_{A1} r_A} \; mod \;\; p$, $R_s = (X'_A)^{1/p_{A1}}$, current session key $K = \left(\frac{X_A}{f k_{AS}(ID_A, ID_B, N_A)}\right)^{r_A R_S} \; mod \; p = g^{r_A r_B r_S} \; mod \;\; p$. A sends the verification message to $\alpha_A = H(ID_A, ID_B, K)$ to B.
  - Same as A, the B computes, $K_{BS} = N_B^{p_{B1}} = g^{p_{B1} r_B} \; mod \;\; p$, $R_s = (X'_B)^{1/p_{B1}}$, current session key$K = \left(\frac{X_B}{f k_{BS}(ID_A, ID_B, N_B)}\right)^{r_B R_S} \; mod \;\; p = g^{r_A r_B r_S} \; mod \;\; p$. B sends the verification message to $\alpha_B = H(ID_A, ID_B, H(K))$ to A.

After verification at both sides, the current session key K is accepted for future communication between A and B.

Here offline password guessing attack is not possible as attacker need to try out every element of set Z_p. With a sufficiently large value of p (say 1024 bit number), the attacker can not break the protocol.

For the impersonate attack, server S generates new R_s every time which is required in session key K. Therefore, compromising one session with every values can not help for other subsequent sessions.

4 Conclusion and Future Work

Key exchange is one of the interesting questions that has taken the attention of many researches. In this notion we have taken the most recent scheme of Loa et al. and given the possible attacks i.e. offline password guessing and impersaction attack. Afterwards, we have modified the existing scheme with minimal changes to maintain the same computation as of Lo et al. However, our scheme requires the sufficient high value of prime order p of a group to maintain the security. In future one may reduce the number of operations and make a scheme without using a hash function.

References

1. Bellovin SM, Merritt M (1992) Encrypted key exchange: password-based protocols secure against dictionary attacks. In: IEEE symposium on security and privacy, pp 72–84, IEEE Computer Society Press
2. Abdalla M, Pointcheval D (2005) Simple password-based encrypted key exchange protocols. In: Menezes A (ed) CT-RSA 2005. LNCS, vol 3376. Springer, Heidelberg, pp 191–208
3. Bellare M, Pointcheval D, Rogaway P (2000) Authenticated key exchange secure against dictionary attacks. In: Preneel B (ed) EUROCRYPT 2000. LNCS, vol 1807. Springer, Heidelberg, pp 139–155
4. Abdalla M, Chevalier C, Pointcheval D (2009) Smooth projective hashing for conditionally extractable commitments. In: Halevi S (ed) CRYPTO 2009. LNCS, vol 5677. Springer, Heidelberg, pp 671–689
5. Boyko V, MacKenzie PD, Patel S (2000) Provably secure password-authenticated key exchange using Diffie-Hellman. In: Preneel B (ed) EUROCRYPT 2000. LNCS, vol 1807. Springer, Heidelberg, pp 156–171
6. Bresson E, Chevassut O, Pointcheval D () Security proofs for an efficient password-based key exchange. In: Jajodia S, Atluri V, Jaeger T (eds) Proceedings of the 10th conference on computer and communications security (ACM CCS 2003), ACM Press, pp 241–250
7. Bresson E, Chevassut O, Pointcheval D (2004) New security results on encrypted key exchange. In: Bao F, Deng R, Zhou J (eds) PKC 2004. LNCS, vol 2947. Springer, Heidelberg, pp 145–158
8. Canetti R, Halevi S, Katz J, Lindell Y, MacKenzie P (2005) Universally composable password-based key exchange. In: Cramer R (ed) EUROCRYPT 2005. LNCS, vol 3494. Springer, Heidelberg, pp 404–421
9. Gennaro R (2008) Faster and shorter password-authenticated key exchange. In: Canetti R (ed) TCC 2008. LNCS, vol 4948. Springer, Heidelberg, pp 589–606
10. Gennaro R, Lindell Y (2003) A framework for password-based authenticated key exchange. In: Biham E (ed) EUROCRYPT 2003. LNCS, vol 2656. Springer, Heidelberg, pp 524–543
11. Katz J, Ostrovsky R, Yung M (2001) Efficient password-authenticated key exchange using human-memorable passwords. In: Pfitzmann B (ed) EUROCRYPT 2001. LNCS, vol 2045. Springer, Heidelberg, pp 475–494
12. Katz J, Vaikuntanathan V (2009) Smooth projective hashing and password-based authenticated key exchange from lattices. In: Matsui M (ed) ASIACRYPT 2009. LNCS, vol 5912. Springer, Heidelberg, pp 636–652
13. Katz J, Vaikuntanathan V (2011) Round-optimal password-based authenticated key exchange. In: Ishai Y (ed) TCC 2011. LNCS, vol 6597. Springer, Heidelberg, pp 293–310
14. Pointcheval D (2012) Exchange password-based authenticated key. PUBLIC KEY CRYPTOGRAPHY - PKC-2012, Lecture notes in computer science, vol 7293. pp 390–397, doi:10.1007/978-3-642-30057-8_23
15. Kobara K, Imai H (2002) Pretty-simple password authenticated key-exchange under standard assumptions. IEICE Trans E85-A(10):2229–2237
16. Bresson E, Chevassut O, Pointcheval D (2004) New security results on encrypted key exchange. In: Proceedings of PKC 2004, LNCS, vol 2947, pp 145–158
17. Boyd C, Montague P, Nguyen K (2001) Elliptic curve based password authenticated key exchange protocols. In: Proceedings of 28th australasian conference on information security and privacy—ACISP 2001, LNCS, vol. 2119, pp 487–501
18. Abdalla M, Pointcheval D (2005) Simple password-based encrypted key exchange protocols. In: Proceedings of topics in cryptology—CT-RSA 2005. LNCS, vol. 3376, pp 191–208

19. Abdalla M, Chevassut O, Pointcheval D (2005) One-time verifier-based encrypted key exchange. In: Proceedings of PKC '05, LNCS, vol. 3386 pp 47–64
20. K. Kobara, H. Imai (2002) Pretty-simple passwordauthenticated key exchange under standard assumptions. IEICE Trans E85-A(10):2229–2237
21. Bellare M, Pointcheval D, Rogaway P (2000) Authenticated key exchange secure against dictionary attacks. In: Proceedings of the advances in cryptology (EUROCRYPT'2000), Springer, Berlin, pp 139–155
22. Bresson E, Chevassut O, Pointcheval D (2004) New security results on encrypted key exchange. In: Proceedings of PKC 2004, LNCS, vol 2947. Springer, Heidelberg, pp 145–158
23. Abdalla M, Pointcheval D (2005) Simple password-based encrypted key exchange protocols. In: Proceedings of topics in cryptology—CT-RSA 2005, LNCS, vol 3376. Springer, Heidelberg, pp 191–208
24. Abdalla M, Chevassut O, Pointcheval D (2005) One-time verifier-based encrypted key exchange. Proceedings of PKC '05, LNCS, vol 3386. Springer, Heidelberg, pp 47–64
25. Ding Y, Horster P (1995) Undetectable on-line password guessing attacks. ACM Oper Syst Rev 29(4):77–86
26. Lin CL, Sun HM, Hwang T (2000) Three party-encrypted key exchange: attacks and a solution. ACM Oper Syst Rev 34(4):12–20
27. Lee TF, Hwang T, Lin CL (2004) Enhanced three-party encrypted key exchange without server public keys. Comput Secur 23(7):571–577
28. Wen HA, Lee TF, Hwang T (2005) Provably secure three-party password-based authenticated key exchange protocol using Weil pairing. IEE Proc Commun 152(2):138–143
29. Nam J, Lee Y, Kim S, Won D (2007) Security weakness in a three-party pairing-based protocol for password authenticated key exchange. Inf Sci 177(6):1364–1375
30. Yeh HT, Sun HM (2004) Password-based user authentication and key distribution protocols for client-server applications. J Syst Softw 72(1):97–103
31. Yoon E-J, Yoo K-Y (2012) Cryptanalysis of an efficient three-party password-based key exchange scheme, In: Procedia Engineering, vol 29, pp 3972–3979, ISSN 1877–7058, doi:10.1016/j.proeng.2012.01.604
32. Steiner M, Tsudik G, Waidner M (1995) Refinement and extension of encrypted key exchange. ACM Oper Syst Rev 29:22–30
33. Lin CL, Sun HM, Hwang T (2000) Three-party encrypted key exchange: attacks and a solution. ACM Oper Syst Rev 34:12–20
34. Chang CC, Chang YF (2004) A novel three-party encrypted key exchange protocol. Comput Stand Interfaces 26(5):472–476
35. Lee TF, Hwang T, Lin CL (2004) Enhanced three-party encrypted key exchange without server public keys. Comput Secur 23(7):571–577
36. Lee SW, Kim HS, Yoo KY (2005) E?cient verifier-based key agreement protocol for three parties without server's public key. Appl Math Comput 167(2):996–1003
37. Sun HM, Chen BC, Hwang T (2005) Secure key agreement protocols for three-party against guessing attacks. J Syst Softw 75:63–68
38. Lu RX, Cao ZF (2007) Simple three-party key exchange protocol. Comput Secur 26:94–97
39. Yoon EJ, Yoo KY (2008) Improving the novel three-party encrypted key exchange protocol. Comput Stand Interfaces 30(5):309–314
40. Phan RCW, Yau WC, Goi BM (2008) Cryptanalysis of simple three-party key exchange protocol (S-3PAKE). Inf Sci 178:2849–2856
41. Guo H, Li Z (2008) Cryptanalysis of simple three-party key exchange protocol. Comput Secur 27:16–21
42. Kim HS, Choi JY (2009) Enhanced password-based simple three-party key exchange protocol. Comput Electr Eng 35:107–114
43. Huang HF (2009) A simple three-party password-based key exchange protocol. Int J Commun Syst 22:857–862
44. Yang JH, Chang CC (2009) An e?cient three-party authenticated key exchange protocol using elliptic curve cryptography for mobile-commerce environments. J Syst Softw 82(9):1497–1502

45. Ding Y, Horster P (1995) Undetectable on-line password guessing attacks. ACM Oper Syst Rev 29(4):77–86
46. Lo NW, Yeh K-H (2010) A practical three-party authenticated key exchange protocol. Int J Innovative Comput Inf Control 6(6):2469–2483

Intrusion Detection in Zero Knowledge System Using Model Checking Approach

Teslin Jacob, Mithun Raman and Sanjay Singh

Abstract The number of services provided to the modern users of todays' generation is countless and ever increasing. However, most of these services require the user to login with a username and a password. These sensitive information has to be sent across the network, which is highly insecure, and can be tapped by any unauthorized individual without much difficulty. The Zero Knowledge protocol provides authentication without the need to send any private and confidential information across the network. Only mathematical computations on these confidential information have to be sent across. In this paper, we have modeled the general working of the Zero Knowledge system by considering the various states that a prover (sender) and a verifier (receiver) will be in during the execution of the protocol, and have proved that the authentication of the prover is possible. Zero Knowledge system is usually considered to be unintrudeable, but that does not stop hackers from attempting to intrude this protocol. So in this paper, we have also considered the various states that an intruder will be in while intruding, and have shown that it is possible for the user to detect if somebody is trying to intrude the Zero Knowledge system. The tool used to model the system is NuSMV.

1 Introduction

Todays generation is exposed to a wide range of services and facilities over the Internet. Most of these services require the user to first register himself; or if he is already registered, then he needs to login, which involves sending of his login credentials over the Internet to the verifier. There are many vulnerabilities and attack vectors for web-based application. This includes both web-specific (i.e. Cross-Site

T. Jacob · M. Raman · S. Singh (✉)
Department of Information and Communication Technology,
Manipal Institute of Technology, Manipal 576104, India
e-mail: sanjay.singh@manipal.edu

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_45,

Scripting [1]), as well as generic (i.e. Password Sniffing [2]), all of which leave the user susceptible to being victims of identity theft. The first technique introduced was the password-based authentication, wherein a user and a host share a password which is essentially a long-term but rather small-size symmetric key. Since in this protocol the password was sent in the open, and the stored password file at the receivers' end was kept in the open which was easily readable by an attacker. To overcome this problem Needham's password protocol [3] was introduced, wherein the receiver uses a one-way function to encode the passwords, which is extremely difficult to invert. But again, the password from the sender to the receiver was sent in the open. So then came the one-time password scheme, and the Encrypted Key Exchange(EKE) protocol [4] which involves adding your own salt.

In all these protocols, there is a common feature, which is sending the user's password over the Internet, either in plaintext format or in an encrypted format. In order to avoid this, the Zero Knowledge system was introduced, which eliminated the need to send confidential information over the Internet [5]. There are still no known attacks on this system. However, hackers can still attempt to attack and break this system. Packets sent across the network can be captured by the intruder to perform cryptanalysis.

The best way to test the different reachable combinations of states of a system is to model it using an appropriate model checking tool. Model checking is one approach where we can test a particular system by representing it in the form of state transition diagrams. The different states of a system and the transitions between them are identified. These can then be modeled as Finite State Machine (FSM). Once this is done, the constraints on the system can be written in the form of logic specifications which should hold true throughout the system's execution in all its states. If it is found that a particular specification is false, then it will return a trace leading to the false condition. Using this trace, it is possible to trace the systems properties and what operations it performs under what conditions. Therefore, using this method, any flaw in the system can be detected, and changes to the model of the system can be made to overcome it. The model checking tool we used in this work is NuSMV (New Symbolic Model Verifier) [6].

In this paper,we have modeled the behavior of the Zero Knowledge system by representing it in terms of the various states that a prover and a verifier will be in, and have checked using Computation Tree Logic (CTL) specifications if authentication is possible. We have also modeled the behavior of an intruder who is trying to capture the packets being sent between the sender and the receiver, and show that it is possible to detect the presence of an intruder.

The remainder of the paper is organized as follows. Sect. 2 discusses about the theoretical background of the Zero Knowledge system. Section 3 describes about the modeling of the Zero Knowledge System in NuSMV. Section 4 discusses the simulation results obtained through NuSMV. Finally conclusion has been drawn in Sect. 5.

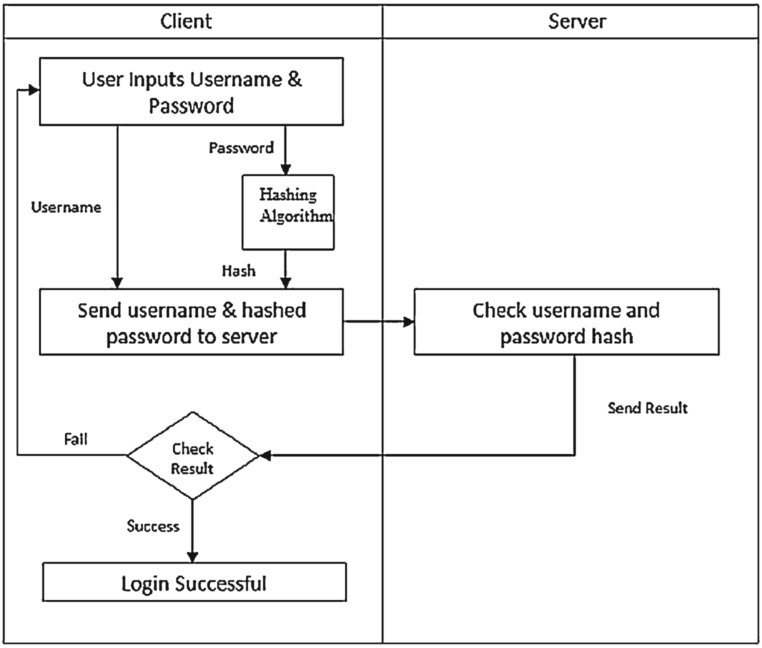

Client
Server
User Inputs Username & Password
Password
Hashing Algorithm
Username
Hash
Send username & hashed password to server
Check username and password hash
Send Result
Fail
Check Result
Success
Login Successful

Fig. 1 Traditional authentication system

2 Theoretical Background

2.1 Current Web Application Login Process

The most common login system used in web application currently is through the use of a form submission of a username and password enabled with SSL communication. In more secure systems, the password is hashed using a Java script-based hashing algorithm before sending it. Figure 1 shows a simple traditional authentication system which shows that an encrypted version of the password needs to be sent across the network.

2.2 Zero Knowledge Proof

Zero Knowledge proofs were first conceived in 1985 in a draft of "The Knowledge Complexity of Interactive Proof-Systems" [7]. While this landmark paper did not invent interactive proof systems, it did invent the **IP** hierarchy of interactive proof systems and conceived the concept of knowledge complexity, a measurement of the amount of knowledge about the proof transferred from the prover to the verifier.

Zero-Knowledge proof is a much popular concept utilized in many cryptography systems. In this concept, two parties are involved, the prover A and the verifier B. Using this technique, it allows prover A to show that he has a credential (for example, a credit card number), without having to give B the exact number. The reason for the use of a Zero-Knowledge Proof in this situation for an authentication system is because it has the following properties [8]:

(i) **Completeness**: if the statement is true, the honest verifier (that is, one following the protocol properly) will be able to prove that the statement is true to an honest verifier every time.
(ii) **Soundness**: if the statement is false, it is not possible (with a very small chance) to fake the result to the verifier that the statement is true.
(iii) **Zero-knowledge**: if the statement is true, the verifier will not know anything other than that the statement is true. Information about the details of the statement will not be revealed.

The step-by-step procedure of the Zero Knowledge password based authentication protocol [8] is given below.

(i) **Initialization**:

- Given group G. Let g_0, g_1 be primitive roots in G.
- Let the public key be $zpk = (\mathrm{G}, g_0)$.

(ii) **Registration Process**:

- User inputs *username* and *password*.
- The user hashes the password with Hash function, H and calculates x = H(password).
- The user then computes $Y = g_0^x$.
- The user sends (*username*, Y) to the server.
- The server stores (*username*, Y) into the database.

(iii) **Authentication Process**:

- The server generates a random one-time token "a" and stores it and sends it to the user.
- User inputs username and password.
- The user hashes the password with Hash function, H and calculates x = H(password).
- The user then computes $Y = g_0^x$.
- The user chooses random $r_x \in \mathrm{G}$ and calculates $T_1 = g_0^{r_x}$.
- The user then calculates $c = H(Y, T_1, a)$ and $z_x = r_x - c_x$.
- The user sends (c, z_x) to the server.
- The server calculates $T_1 = Y^c g_0^{z_x}$ and verifies that $c = H(Y, T_1, a)$.
- If successful, user is authenticated.

The above algorithm is based on a non-interactive sigma protocol [9], which is a technique commonly used to prove the knowledge of a variable, which in this case, is the password. The user who is logging in is the prover, and the server verifying the login is the verifier.

Here are some of the explanation of the components in the algorithm:

- G: This is a cyclic group. This group contains a set of numbers which is based on a formula. This is a public group which will be available to both prover (user) and verifier (server).
- g_0: A generator of the group G. It is an element of the group G. This is a public variable which will be available to both prover (user) and verifier (server).
- x: The hash of the password that the user inputs.
- Y: The pseudonym of the user. This is used for the verifier in the calculation of the proof of knowledge.
- a: The random token generated for each login attempt.
- T_1, r_x, z_x, c: Other miscellaneous variables which are used in the calculation.

2.3 Introduction to NuSMV

For modeling and verification of properties of Zero Knowledge system we have used NuSMV. NuSMV is a reimplementation and extension of SMV symbolic model checker, the first model checking tool based on Binary Decision Diagrams (BDDs). The tool has been designed as an open architecture for model checking. It is aimed at reliable verification of industrial sized designs, for use as a back end for other verification tools and as a research tool for formal verification techniques [10].

NuSMV supports the analysis of specifications expressed in temporal logic. User interaction is performed with a textual interface, as well as in batch mode. The system is represented in terms of a set of variables, predicates and transitions. Once this is done the constraints on the system can be represented in the form of Computation Tree Logic (CTL) or Linear-time Temporal Logic (LTL) [11] specifications which should hold true throughout the systems execution in all its states. Whenever a specification is false, the system returns a counterexample to the user indicating the sequence of steps that led to the counterexample. Using these counterexamples, it is possible to trace the systems properties and what operations it performs under what conditions. Accordingly, any flaw in the system can be detected and changes to the model of the system can be made to overcome it.

3 Modeling Zero Knowledge Protocol in NuSMV

As stated in Sect. 2, the Zero Knowledge protocol revolves around two major entities, the Prover (sender), and the verifier (receiver). The prover tries to prove a fact to the verifier without actually handing over the fact across the network to the receiver. Here

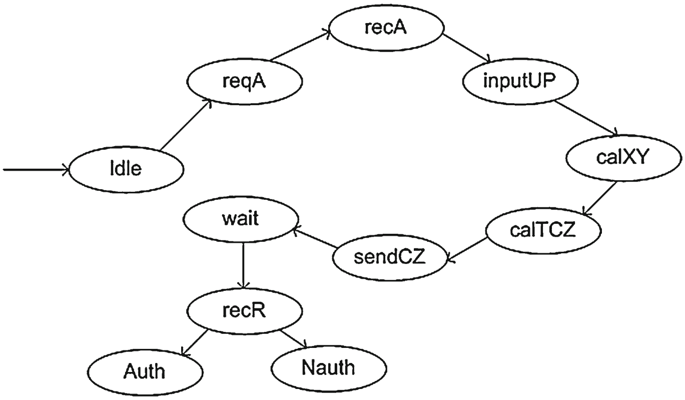


Fig. 2 Various states of a prover

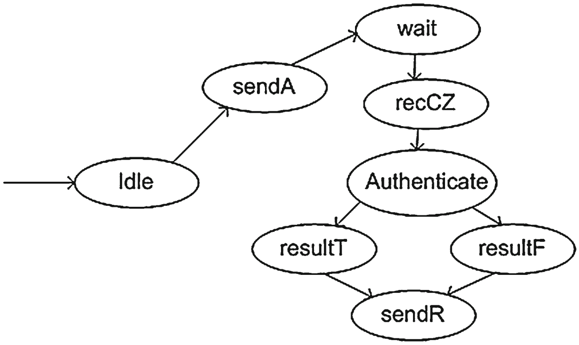


Fig. 3 Various states of a verifier

we have modeled the various states of the prover and the verifier while executing the Zero Knowledge protocol. Figure 2 shows the various states that a prover undergoes through, while requesting for authentication from the verifier.

Figure 3 shows the various states that a verifier goes through on an authentication request from the prover.

Using NuSMV, we have modeled the FSM shown in Figs. 2 and 3 respectively, and have considered the following cases (specifications) to verify the functionality of Zero Knowledge systems:

Case I: When the verifier verifies that the prover is authentic, the prover will move into a state where he is authenticated.

Case II: If the prover is not authentic, and the verifier detects that the prover is not authentic, the prover moves into a state where he is declared that he is not authenticated.

Case III: If the verifier declares him to be non-authentic, then the prover cannot move into an authenticated state.

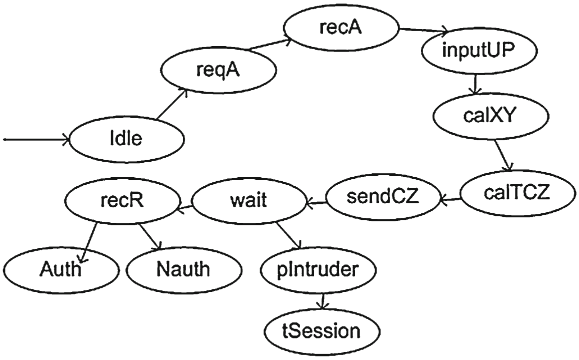


Fig. 4 Various states of a prover in the presence of an intruder

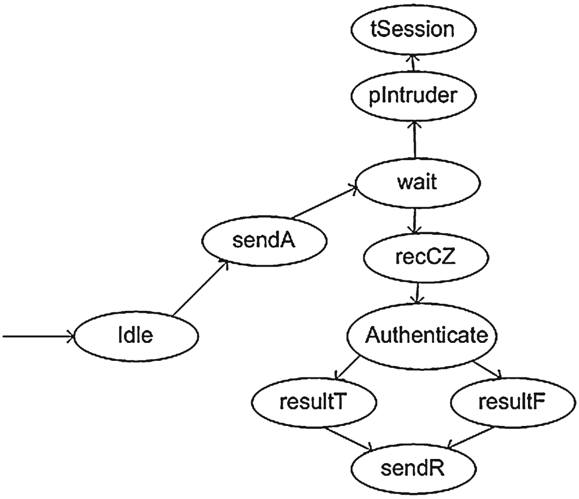


Fig. 5 Various states of a verifier in the presence of an intruder

Case IV: If the verifier declares him to be authentic, then the prover cannot move into a non-authenticated state.

So far we have explained the working of a Zero Knowledge system as an authentication protocol between two users, prover and verifier. Now we take into consideration the presence of an intruder in the system. The above FSMs are not equipped with special states that can be used to detect an intruder. Therefore, we present the following two modified FSMs in Figs. 4 and 5 that show the states in the presence of an intruder.

Figure 6 shows the FSM of an intruder who is responsible for launching an attack on the protocol. The additional state (*pIntruder*) helps to detect intrusion. After the prover or the verifier have sent some message across the network, their respective

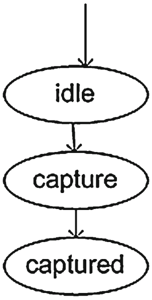
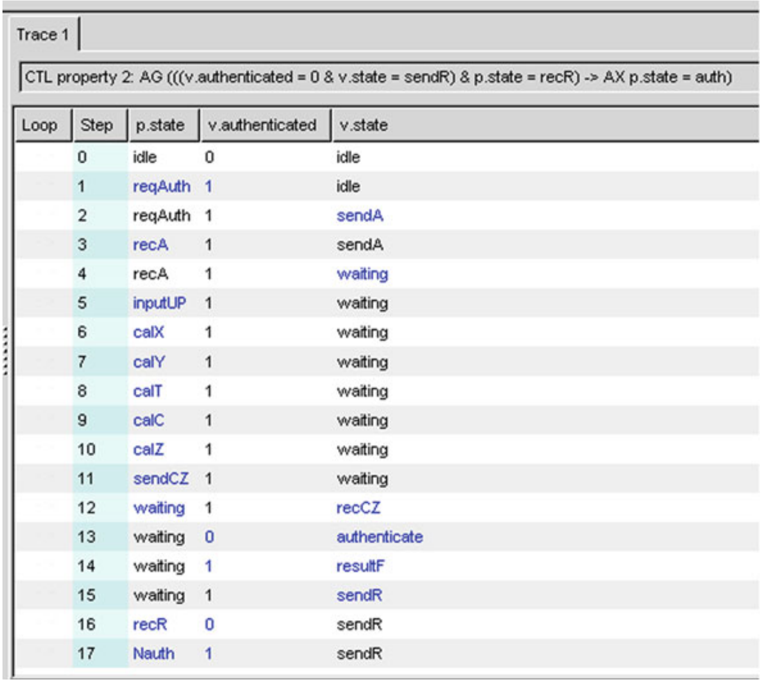


Fig. 6 Various states of an intruder



CTL property 2: AG (((v.authenticated = 0 & v.state = sendR) & p.state = recR) -> AX p.state = auth)

Loop	Step	p.state	v.authenticated	v.state
	0	idle	0	idle
	1	reqAuth	1	idle
	2	reqAuth	1	sendA
	3	recA	1	sendA
	4	recA	1	waiting
	5	inputUP	1	waiting
	6	calX	1	waiting
	7	calY	1	waiting
	8	calT	1	waiting
	9	calC	1	waiting
	10	calZ	1	waiting
	11	sendCZ	1	waiting
	12	waiting	1	recCZ
	13	waiting	0	authenticate
	14	waiting	1	resultF
	15	waiting	1	sendR
	16	recR	0	sendR
	17	Nauth	1	sendR

Fig. 7 Trace of case III

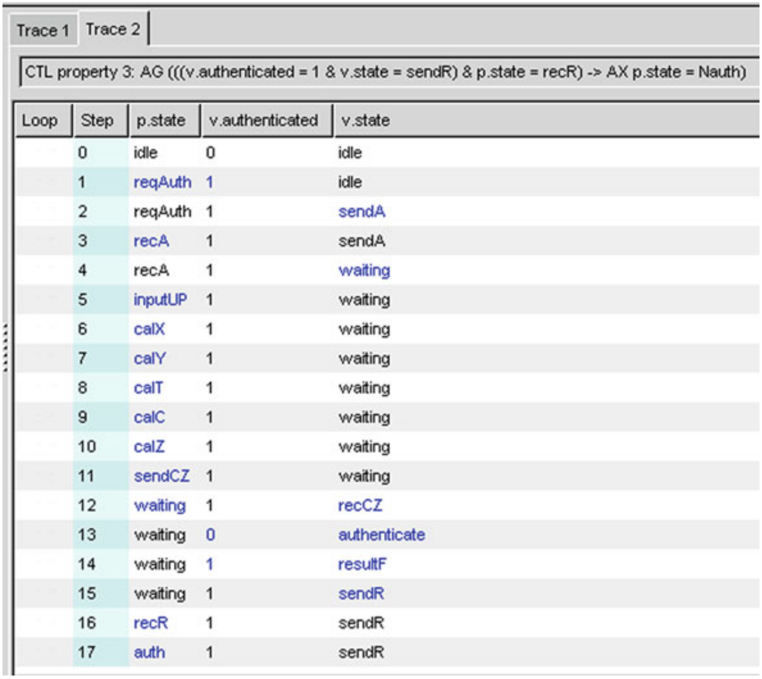
counters will start ticking until they receive a response from the opposite party. If this tick count value exceeds some threshold, the parties can conclude that either an intruder has captured the messages, or the messages are lost due to transmission error. If such a case arises, the parties will move into the *pIntruder* state and finally enter the *tSession* state where the session is terminated. These results are shown in Sect. 4.

4 Simulation Results and Discussion

We first present the results for the modeling of the classical Zero Knowledge system. The first case (i.e specification) of Sect. 3 can be represented in NuSMV using the following specification:
AG (((v.authenticated = 1 & v.state = sendR) & p.state = recR) → AX p.state = auth).
The result obtained was TRUE which is indeed correct.

The second case in Sect. 3 can be represented in NuSMV using the following specification:
AG (((v.authenticated = 0 & v.state = sendR) & p.state = recR) → AX p.state = Nauth).
The result obtained was TRUE which is correct because the prover should go into a non-authenticated state when the verifier declares that he is unauthentic.

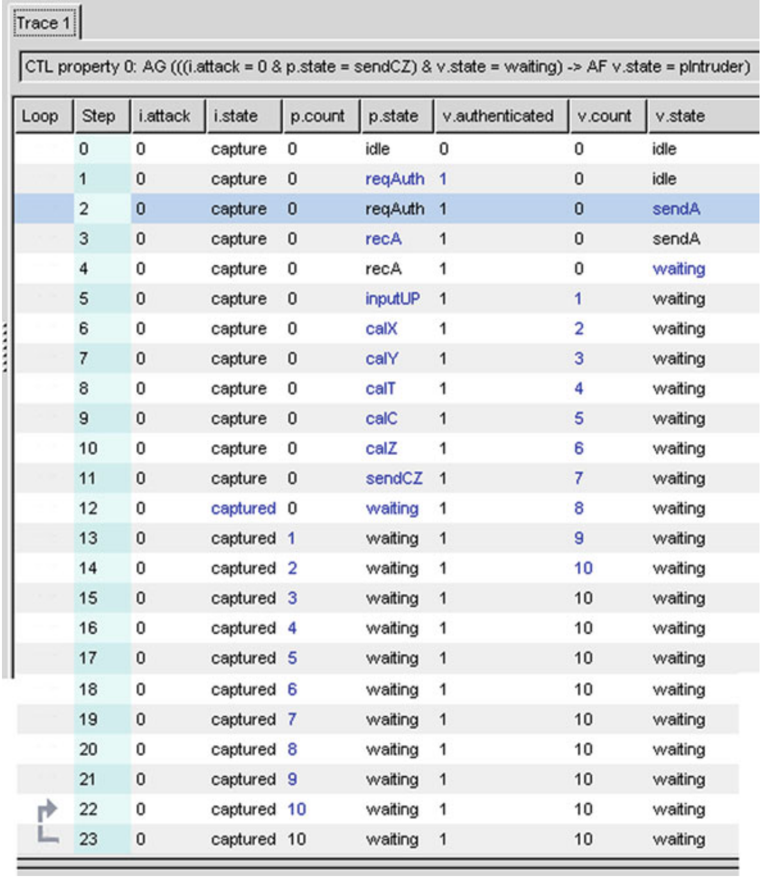
The third case in Sect. 3 can be represented in NuSMV using the following specification:
AG (((v.authenticated = 0 & v.state = sendR) & p.state = recR) → AX p.state = auth).
The result obtained was FALSE which is correct because the prover cannot go into an authenticated state when the verifier declares that he is unauthentic. The trace



CTL property 3: AG (((v.authenticated = 1 & v.state = sendR) & p.state = recR) -> AX p.state = Nauth)

Loop	Step	p.state	v.authenticated	v.state
	0	idle	0	idle
	1	reqAuth	1	idle
	2	reqAuth	1	sendA
	3	recA	1	sendA
	4	recA	1	waiting
	5	inputUP	1	waiting
	6	calX	1	waiting
	7	calY	1	waiting
	8	calT	1	waiting
	9	calC	1	waiting
	10	calZ	1	waiting
	11	sendCZ	1	waiting
	12	waiting	1	recCZ
	13	waiting	0	authenticate
	14	waiting	1	resultF
	15	waiting	1	sendR
	16	recR	1	sendR
	17	auth	1	sendR

Fig. 8 Trace of case IV



CTL property 0: AG (((i.attack = 0 & p.state = sendCZ) & v.state = waiting) -> AF v.state = pIntruder)

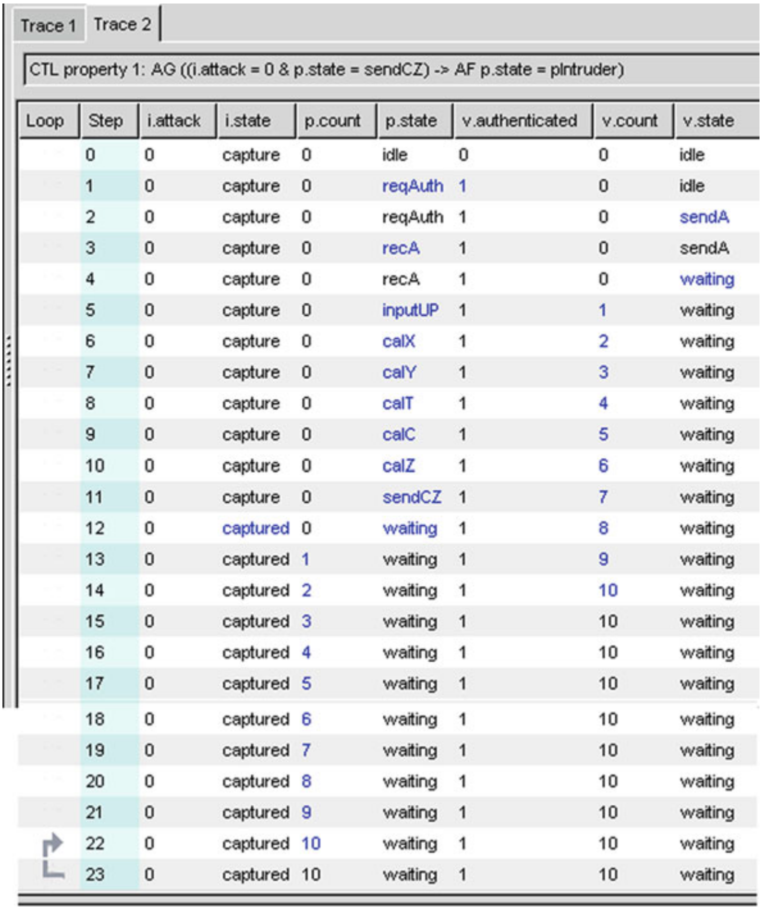
Loop	Step	i.attack	i.state	p.count	p.state	v.authenticated	v.count	v.state
	0	0	capture	0	idle	0	0	idle
	1	0	capture	0	reqAuth	1	0	idle
	2	0	capture	0	reqAuth	1	0	sendA
	3	0	capture	0	recA	1	0	sendA
	4	0	capture	0	recA	1	0	waiting
	5	0	capture	0	inputUP	1	1	waiting
	6	0	capture	0	calX	1	2	waiting
	7	0	capture	0	calY	1	3	waiting
	8	0	capture	0	calT	1	4	waiting
	9	0	capture	0	calC	1	5	waiting
	10	0	capture	0	calZ	1	6	waiting
	11	0	capture	0	sendCZ	1	7	waiting
	12	0	captured	0	waiting	1	8	waiting
	13	0	captured	1	waiting	1	9	waiting
	14	0	captured	2	waiting	1	10	waiting
	15	0	captured	3	waiting	1	10	waiting
	16	0	captured	4	waiting	1	10	waiting
	17	0	captured	5	waiting	1	10	waiting
	18	0	captured	6	waiting	1	10	waiting
	19	0	captured	7	waiting	1	10	waiting
	20	0	captured	8	waiting	1	10	waiting
	21	0	captured	9	waiting	1	10	waiting
↱	22	0	captured	10	waiting	1	10	waiting
↳	23	0	captured	10	waiting	1	10	waiting

Fig. 9 Trace showing first specification for intruder

obtained is shown in Fig. 7 which shows the sequence of steps which proves that the specification is wrong.

The forth case in Sect. 3 can be represented in NuSMV using the following specification:
AG (((v.authenticated = 1 & v.state = sendR) & p.state = recR) → AX p.state = Nauth).
The result obtained was FALSE which is correct because the prover should go into an authenticated state when the verifier declares that he is authentic. The trace obtained is shown in Fig. 8 which shows the sequence of steps which proves that the specification is wrong.

Next we present the results for the modeling of the intrusion detection states.



CTL property 1: AG ((i.attack = 0 & p.state = sendCZ) -> AF p.state = pIntruder)

Loop	Step	i.attack	i.state	p.count	p.state	v.authenticated	v.count	v.state
	0	0	capture	0	idle	0	0	idle
	1	0	capture	0	reqAuth	1	0	idle
	2	0	capture	0	reqAuth	1	0	sendA
	3	0	capture	0	recA	1	0	sendA
	4	0	capture	0	recA	1	0	waiting
	5	0	capture	0	inputUP	1	1	waiting
	6	0	capture	0	calX	1	2	waiting
	7	0	capture	0	calY	1	3	waiting
	8	0	capture	0	calT	1	4	waiting
	9	0	capture	0	calC	1	5	waiting
	10	0	capture	0	calZ	1	6	waiting
	11	0	capture	0	sendCZ	1	7	waiting
	12	0	captured	0	waiting	1	8	waiting
	13	0	captured	1	waiting	1	9	waiting
	14	0	captured	2	waiting	1	10	waiting
	15	0	captured	3	waiting	1	10	waiting
	16	0	captured	4	waiting	1	10	waiting
	17	0	captured	5	waiting	1	10	waiting
	18	0	captured	6	waiting	1	10	waiting
	19	0	captured	7	waiting	1	10	waiting
	20	0	captured	8	waiting	1	10	waiting
	21	0	captured	9	waiting	1	10	waiting
↱	22	0	captured	10	waiting	1	10	waiting
└	23	0	captured	10	waiting	1	10	waiting

Fig. 10 Trace showing second specification for intruder

If the prover is in the sending state, and the intruder has attacked, and the verifier is waiting for prover's message, then the verifier should go into the *pIntruder* state. This is represented by the following specification:
AG (((i.attack = 1 & p.state = sendCZ) & v.state = waiting) → AF v.state = pIntruder).
The result obtained was TRUE which is correct.

If the prover has sent a message, and the intruder has captured the message, then the prover should move into the *pIntruder* state. This is represented by the following specification:
AG ((i.attack = 1 & p.state = sendCZ) → AF p.state = pIntruder).
The result obtained was TRUE which is correct.

Now if the prover has sent a message which was not captured by the intruder, and the verifier is waiting for the message, then the verifier cannot go into the *pIntruder* state. The following specification shows this:
AG (((i.attack = 0 & p.state = sendCZ) & v.state = waiting) → AF v.state = pIntruder).
The result returned was indeed FALSE, with Fig. 9 showing its trace.

If the prover has sent a message which was not captured by the intruder, then the prover cannot go into the *pIntruder* state. The following specification shows this: AG ((i.attack = 0 & p.state = sendCZ) → AF p.state = pIntruder).
The result returned was indeed FALSE, with Fig. 10 showing its trace.

5 Conclusion

In this paper through model checking approach we have shown that while using Zero Knowledge System, it is possible for a prover to be authenticated if the verifier is convinced so. We have also shown that it is possible for both the prover and the verifier to detect if an intruder is listening to their conversation, and so can terminate their session to prevent further harm from the intruder. Other aspects of the protocol can be modeled by taking into consideration other possible kinds of attacks, and by step-by-step tracing of the execution path, any other flaws and improvements can be detected, and solutions can be suggested using model checking. Hence, model checking provides a very essential tool to verify all kinds of systems that can be represented as state transition diagrams, especially for verifying the various security protocols for their strengths and weaknesses, and suggest improvements which can lead to better system designs and improved security and quality.

References

1. Wassernann G, Davis C (2008) Static detection of cross-site scripting vulnerabilities. In: Software engineering, 2008. ICSE ’08 ACM/IEEE 30th international conference, pp 171–180
2. Trabelsi Z, Rahmani H, Frikha M (2004) Malicious sniffing systems detection platform. In: Applications and the internet, 2004. Proceedings. 2004 international symposium, pp 171–180
3. Sultana S, Jabiullah M, Rahman M (2009) Improved needham-schroeder protocol for secured and efficient key distributions. In: Computers and information technology, 2009. ICCIT ’09. 12th international conference, pp 564–569
4. Barmawi A, Takada S, Doi N (1997) Augmented encrypted key exchange using rsa encryption. In: Personal, indoor and mobile radio communications, 1997. Waves of the year 2000. PIMRC ’97. The 8th IEEE international symposium, pp 490–494
5. Guilou LC, Berson TA (1990) How to explain zero-knowlege protocols to your children. In: Advances in cryptology—CRYPTO ’89, pp 628–631
6. FBK-IRST: Nusmv. (2012) http://nusmv.fbk.eu/NuSMV/index.html
7. Goldwasser S, Micali S, Rackoff C (1989) The knowledge complexity of interactive proof systems. SIAM J Comput 18:186–208

8. Jun LJ (2010) Brandon: implementing zero-knowledge authentication with zero knowledge. In: The Python papers monograph 2:9 proceedings of PyCon Asia-Pacific
9. Barak B (2010) Zero knowledge, identification protocols. www.cs.princeton.edu/courses/archive/spr10/cos433/lec18new
10. Cavada R, Cimatti A, Jochim CA, Keighren G, Olivetti E, Pistore M, Roveri M, Tchaltsev A (2010) Nusmv 2.5 user manual. http://nusmv.fbk.eu/NuSMV/userman/index-v2.html
11. Huth M, Ryan M (2004) Logic in computer science: modelling and reasoning about systems. Cambridge University Press, New York

Detecting Malicious Users in P2P Streaming Systems by Using Feedback Correlations

Feng-Li Zhang, Yang Bai, Jie Hou and Yuan-Wei Tan

Abstract The trust and reputation models were introduced to restrain the impacts caused by rational but selfish peers in P2P streaming systems. However, these models face with two major challenges from dishonest feedback and strategic altering behaviors. To answer these challenges, we present a global trust model based on network community, evaluation correlations, and punishment mechanism. We also propose a two-layered overlay to provide the function of peers behaviors collection and malicious detection. The simulation results show that our trust framework can successfully filter out dishonest feedbacks by using correlation coefficients.

1 Introduction

With the increasing popularity of P2P steaming systems, many security issues have to be faced by both end users and service providers. While Worms, Trojans, and viruses as the most typical problems threatening P2P streaming systems, P2P streaming facing many new problems. On one hand, because piracy can be much easier distributed in the network, intellectual property protection (IPR) problems is noteworthy; On the other hand, attacks utilizing P2P streaming system features are the potentially devastating security issues, such as free-riding, data pollution, routing attacks, and index poisoning. As an effective way to solve these problems, trust models have been proposed [1–4]. As one of typical models, Eigentrust [2] evaluates a peers global trust value based on its history upload-behaviors, with which a peer can reject low value peers to avoid getting bad quality services. But this model did not consider strategic

F.-L. Zhang · Y. Bai (✉) · J. Hou · Y.-W. Tan
Department of Computer Science and Engineering, University of Electronic and Science Technology of China, Chengdu, China
e-mail: baiyang.cncq@gmail.com

F.-L. Zhang
e-mail: fzhang@uestc.edu.cn

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_46,

altering behaviors. Consequently, Chang et al. proposed the DyTrust [5] model with punishment schemes to reduce the affection of strategic altering behaviors.

In this paper, we propose a two-layered overlay to provide the function of peers behaviors collection and malicious user detection. The lower level of the overlay is organized as a mesh by streaming peers, which is responsible for medium downloading and sharing. The upper level of the overlay is based on KAD [6], which stores the history evaluation of the peers and calculates their state-of-art trust values. We also propose a trust framework integrating service objects, network community, and trust value correlations. A series experiments have been carried out to analyze the performance of our model. The evaluating results show that our model can effectively filter out dishonest feedbacks, detect malicious behaviors and resist altering behaviors.

2 Related Work

In feedback-evaluation based models, users evaluate each service and calculate the trust value based on the evaluations. Some of Feedback-evaluation based models provide global trust value, others provide local trust value. P2Prep proposed a reputation sharing protocol for P2P file sharing systems [7], in which node keeps track and shares reputation with others. P2Prep uses polling algorithm to ensure nodes anonymity in streaming sharing process. Bayesian network-based trust model provides a flexible method to represent differentiated trust in aspects of each others capability and combine different aspects of trust. With the limitation of Bayesian network, the computing cost rises with the size of the network. As the consequence, this model is only applicable in relative small-size networks. DyTrust [5] model was proposed with a punishment schemes, and can restrain the affection of strategic altering behaviors. These models can decrease both network expense and trust-building delay. Eigentrust [2], peerTrust [4] are global trust provided models. Eigentrust [2], based on service providers history upload-behaviors, calculates out a global trust value for each node, with which low quality service downloading can be decreased. In [4], peerTrust presented a coherent adaptive trust model for quantifying and comparing the trustworthiness of peers based on a transaction-based feedback mechanism. These models have effect to defense with free-riding and cooperative attacks.

3 Trust Framework

3.1 Overlay Structure

The upper level of the overlay is based on KAD, while the lower level of the overlay is structured as a mesh. Peers of the upper level (namely upper nodes) gather and store the evaluating information. The upper nodes are also responsible to the trust value calculating and retrieving. The lower level of the overlay is organized as a mesh by streaming peers, which is responsible for medium downloading and sharing. When a peer of the lower level (namely lower node) joins into the system, it chooses several lower nodes as its neighbors randomly (Fig. 1).

3.2 Trust Model and Computation

We assume the lower nodes get streaming data and report their evaluations in certain duration $[t_{start}, t_{end}]$, we divide it into several sub segments which are called timeframes.

(1) **The direct evaluation:** The direct evaluation is the average of several evaluation values between two fixed nodes with a same resource. It is the original data based on feedback. In our model, i represents the lower node requesting streaming data, j represents the lower node responding with the data, R denotes the requested service. m stands for the frequency that i getting R from j, in the nth timeframe; let $e_{ij}^{n}(R)$ denote the estimated value i to R in node j; the direct evaluation from i to R of j, $D_{ij}^{n}(R)$ is defined as formula (1)

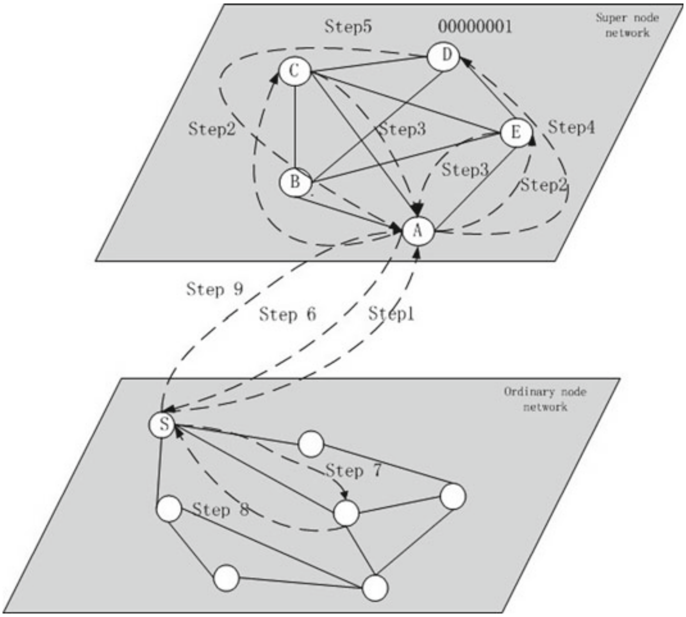


Fig. 1 Communication process of the trust framework

$$D_{ij}^{n}(R)=\frac{1}{m}\sum e_{ij}^{n}(R). \tag{1}$$

(2) **The expected direct evaluation value:** The direct evaluation is the average of several evaluation values between several nodes with a same resource. It reflects the resources general situation. In nth timeframe, node b denotes one of user set $clt(j)$, which gets R from node j, the element number of set $clt(j)$ is denoted by m. The expected value of Rs directly evaluation is defined as formula (2)

$$D_{j}^{n}(R)=\frac{1}{m}\sum_{b\in Clt(j)} D_{bj}^{n}(R). \tag{2}$$

(3) **The correlation coefficient based filter function:** The correlation coefficient based filter function utilize Pearson correlation coefficient to metric the correlation ship between current nodes direct evaluation and the expected direct evaluation value. Let $[D_{ij}^{1}(R)\cdots D_{ij}^{t}(R)\cdots D_{ij}^{n}(R)]$ denote the direct evaluation vector from i to R of j, from the 1th timeframe to the nth timeframe. Let $[D_{j}^{1}(R)\cdots D_{j}^{t}(R)\cdots D_{j}^{n}(R)]$ denote the expected value of all lower nodes direct evaluation vector, each element of the direct evaluation vector can be calculated with formula (2). We use the Pearson correlation coefficient to measure the correlation between the vector $[D_{ij}^{1}(R)\cdots D_{ij}^{t}(R)\cdots D_{ij}^{n}(R)]$ and $[D_{j}^{1}(R)\cdots D_{j}^{t}(R)\cdots D_{j}^{n}(R)]$, which is defined as formula (3)

$$\begin{aligned} corr_{ij}^{n}(R)=&\frac{1}{n-1}\sum_{t=1}^{n}\left(\frac{D_{ij}^{t}(R)-\frac{1}{n}\sum_{k=1}^{n}D_{ij}^{k}(R)}{\sigma_{D_{ij}^{t}(R)}}\right)\\ &\times\left(\frac{D_{j}^{t}(R)-\frac{1}{n}\sum_{k=1}^{n}D_{j}^{k}(R)}{\sigma_{D_{j}^{t}(R)}}\right). \end{aligned} \tag{3}$$

In (3), n denotes the current timeframe number, let $\sigma_{D_{ij}^{t}(R)}$ and $\sigma_{D_{j}^{t}(R)}$ denote respectively the standard deviations of $[D_{ij}^{1}(R)\cdots D_{ij}^{t}(R)\cdots D_{ij}^{n}(R)]$ and $[D_{j}^{1}(R)\cdots D_{j}^{t}(R)\cdots D_{j}^{n}(R)]$. According to Pearson correlation coefficient, if $corr_{ij}^{n}\geq$ *corrThreshold*, it indicates that the relation between $[D_{ij}^{1}(R)\cdots D_{ij}^{t}(R)\cdots D_{ij}^{n}(R)]$ and $[D_{j}^{1}(R)\cdots D_{j}^{t}(R)\cdots D_{j}^{n}(R)]$ is acceptable; $corr_{ij}^{n} < corrThreshold$ manifests that the relation between $[D_{ij}^{1}(R)\cdots D_{ij}^{t}(R)\cdots D_{ij}^{n}(R)]$ and $[D_{j}^{1}(R)\cdots D_{j}^{t}(R)\cdots D_{j}^{n}(R)]$ have large discrepancy. Namely, the users direct evaluation is not in accordance with the expected one, which may mean that the user is not honest. As the result, the model keeps suspicious attitude to the evaluations, and will filter out these dishonesty evaluations by using the $flag_{ij}^{n}(R)$. The filtering mechanism is defined as formula (4), (5) and (6)

$$flag_{ij}^{n}(R) = \begin{cases} 0, \; corr_{ij}^{n}(R) < corrThreshold_{j}^{n}(R); \\ 1, \; else. \end{cases} \tag{4}$$

$$corrThreshold_{j}^{n}(R) = \frac{1}{k} \sum_{i \in clt(j)} corr_{ij}^{n}. \tag{5}$$

(4) **The accumulation feedback quality:** In formula (6), b is a lower node receiving streaming data from j, Cu_{ib}^{n} denotes the credibility from node b to i, and Q_{bj}^{n} denotes the accumulation feedback quality of service R of j during the nth timeframe

$$Q_{ij}^{n}(R) = \frac{\sum_{b \in Clt(j)} Cu_{ib}^{n}(D_{bj}^{n}(R))}{\sum_{b \in Clt(j)} flag_{bj}^{n}(R) Cu_{ib}^{n}}. \tag{6}$$

(5) **The current trust value:** The current trust value defines the integration of history trust value and the accumulation feedback quality of a service. It reflects the peers behavior and the quality of service that provided. Let $S_{j}^{n}(R)$ denotes the current trust value of service R of j, which can be defined as formula (7)

$$S_{j}^{n}(R) = (1 - \rho) S_{j}^{n-1}(R) + \rho Q_{j}^{n}(R). \tag{7}$$

(6) **The group trust:** The group trust reflects the influence degree form peers location group to other groups. Considering clustering property of P2P streaming networks, we define the group trust value T_{g}^{n} in the nth timeframe as (8), which depends on the trust values between nodes inside and outside the group

$$T_{g}^{n} = \frac{1}{k} \sum_{j \in e} \frac{1}{r} \sum_{l \notin g} D_{ij}^{n} \times \frac{2E_j}{k_j(k_j - 1)}, \tag{8}$$

where k denotes the number of edge nodes in g, and e denotes the set of edge nodes in a group g. Let j denote one of the edge nodes. l stands for a set of nodes outside of the group g connecting with j. The parameter r is the node number in the set e, where k_j represents the number of connected nodes of node j in g; and E_j is the real connected edge number between these nodes. $\frac{2E_j}{k_j(k_j-1)}$ denotes the clustering coefficient of node j in group g. In a group, if the clustering coefficient of a node is higher, the influence of its trust value between group members would be larger.

(7) **The global trust:** We define the global trust is constituted by the current trust value and the group trust value. So that we consider the resources trust not only with its QoS, the providers behavior, but also the influence degree. Let θ be the weighted factor of trust value. The global trust of R of j can be described as:

$$GT_{j}^{n}(R) = \theta \times S_{j}^{n}(R) + (1 - \theta) T_{g}^{n}. \tag{9}$$

4 Experimental Evaluation

4.1 Convergence

Figure 2 illustrates the trust building and milking process, which shows that the malicious user with initial trust value being 1.0 will lose 0.9 trust value only after 7 timeframes, while it costs about 20 timeframes to build up a users trust value from 0.1 to 0.8. It indicates that the cost of building up ones reputation is much higher than that of milking it. On the other hand, when a users trust value becomes relatively high or low, namely 0.92 or 0.03 in Fig. 2, it may get steady, which shows that the model can achieve a convergence state. The results indicate that out trust model can successfully distinguish malicious behavior and good behavior in limited timeframes.

4.2 Sensitiveness to Strategically Altering Behaviors

Users may strategically change their behaviors to attack the trust model. An attacker can repeatedly behave well to increase their trust values, while carry out malicious behaviors when their trust values are high enough. To test the sensitiveness of our trust model in the experiments, we let the user behave well when its trust value is lower than 0.05, and be malicious once its trust value exceeds 0.85.

Figure 3 shows the variations of the trust values with the strategically altering behaviors. The line with circles means the current user behavior status. When it stays at 1.0, it shows that the user performs well. While the line stays at 0, it means

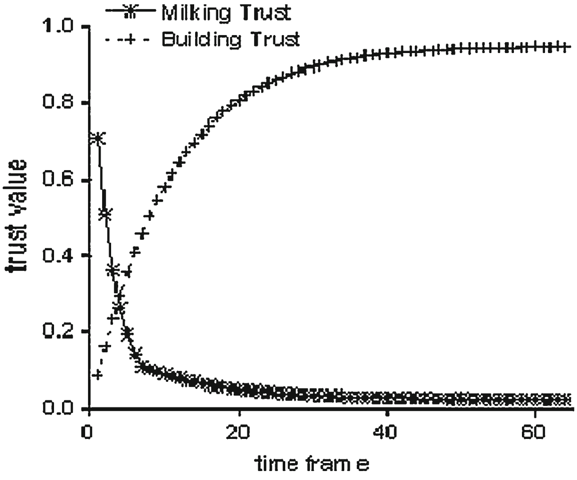


Fig. 2 Building trust versus milking trust

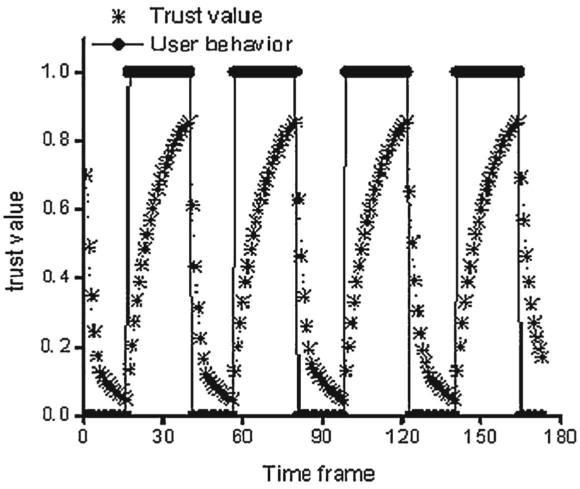


Fig. 3 Strategically altering behaviors

the user is behaving maliciously. The other line with stars represents the users real trust value. The results show that it takes about 11 time frames to increase the users trust value from 0.2 to 0.6 which is much more than that it takes to decrease from 0.6 to 0.2. We can conclude that our trust model is sensitive to realize strategically altering behaviors, and can defend against the threat effectively.

4.3 Effectiveness Against Dishonest Feedback

A dishonest user may provide fake evaluations. One way is to spamming, which means an attacker gives out random evaluations. The other is to speaking irony, that an attacker provides adverse evaluations. The dishonest feedback may cause the system evaluations fail to reflect its real situations.

Figure 4 shows the FNR for different dishonest rate ranging from 0.0 to 0.5, the increasing step is 0.1. The results show that the FNR drops directly from the initial largest value 0.95 to zero within few timeframes for each dishonest rate. It is very interesting that the FPRs in the experiments were always zero. These results indicate that our trust model can effectively detect malicious users with the existence of dishonest feedbacks in relatively short time.

Figure 5 illustrates the FNR for different malice rate ranging from 0.1 to 0.4. The results show that the FNR converges to near zero within few timeframes, which indicate that the trust model can successfully detect malicious users with the existence of dishonest feedback, even under the condition that the malice ratio is 0.4.

Figure 6 represents the detection FNR variation of two models. Figure 7 represents the detection FPR variation of two models. The result denotes that our model have

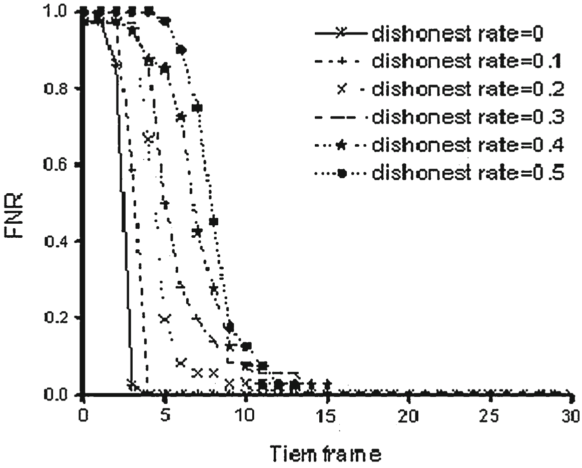


Fig. 4 The FNR for different dishonest rate of the trust model (malice rate = 0.3)

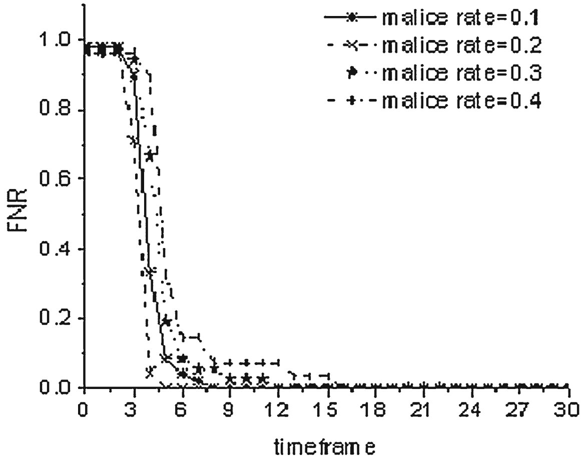


Fig. 5 The FNR for different malice rate of the trust model (dishonest rate = 0.2)

high FNR value at the begin, but then decrease to 0, while DyTrust keep at a stable FNR value at about 0.4. the FPR of our model keep at 0, while the DyTrust model keeps at 0.6 for about 60 cycles. The main reason behind this can be concluded as: after a short initial time, the vector of direct evaluation can reflect service providers real situation; so that the Pearson correlation coefficient works well to filter out the dishonest evaluation, hence, the final trust value can represents peers real situation. In this way, the accuracy of malicious detection can be improved obviously. In conclusion our model has more accuracy detection efficiency than DyTrust model.

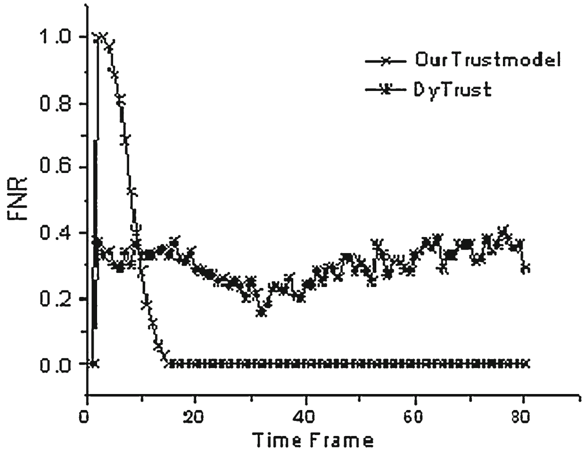


Fig. 6 The FNR comparison between our model and DyTrust model (dishonest rate = 0.2, malice rate = 0.4)

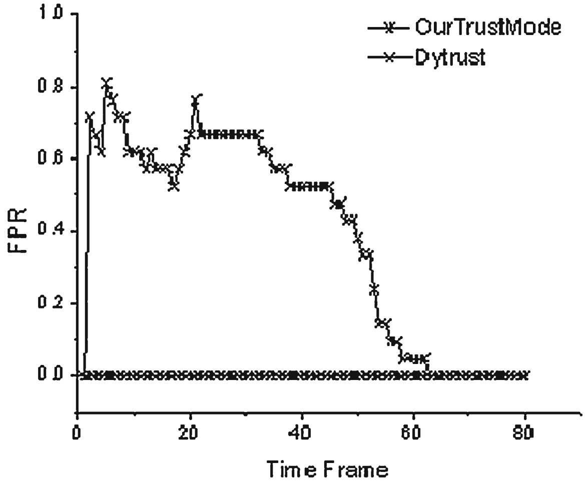


Fig. 7 The FPR comparison between our model and DyTrust model (dishonest rate = 0.2, malice rate = 0.4)

5 Conclusion and Future Work

In this paper, we proposed a novel trust framework in distributed streaming systems. We introduced the evaluation feedback, history trust value, and group trust value into the calculation of a global trust value. The correlation coefficient detecting method is involved to filter out dishonest evaluation feedback, and the punishment mechanism is used to defend against strategic altering behaviors attacks. The simulation results show that our trust framework can effectively detect malicious users within around 10 timeframes, and can resist dishonest feedback attack effectively. In the future, we will consider the scoring system of feedback evaluation based on streaming system details.

Acknowledgments We thank the anonymous reviewers for their helpful comments. This work is supported by a SafeNet Research Award, by key technology research and development program of Sichuan province under grant $No.\ 2010FZ0101$, and by important national science & technology specific projects under grant $No.\ 2011ZX03002$-002-03.

References

1. Feng J, Zhang Y, Wang H (2010) A trust management model based on bi-evaluation in p2p networks. IEICE Trans Inf Syst E93D:466–472
2. Kamvar S, Schlosser M, Garcia-Molina H (2003) The EigenTrust algorithm for reputation management in P2P networks. ACM Press, Budapest
3. Luo J, Ni X (2009) A clustering analysis and agent-based trust model in a grid environment supporting virtual organizations. Int J Web Grid Services 5(1):3–16
4. Xiong L, Liu L (2004) PeerTrust: supporting reputation-based trust for peer-to-peer electronic communities. IEEE Trans Knowl Data Eng 16(7):843–857
5. Chang J, Wang H, Gang Y (2006) A dynamic trust metric for p2p system. In: Proceedings of the 5th international conference of grid and cooperative computing workshops, Changsha, China, pp 117–120
6. Locher T, Schmid S, Wattenhofer R (2011) eDonkey & eMule's Kad: measurements & attacks. Fundam Inf 109:383–403
7. Selcuk AA, Uzun E, Pariente MR (2008) A reputation-based trust management system for p2p networks. Int J Netw Secur 6(3):235–245
8. Yao W, Julita V (2003) Bayesian network-based trust model. Web intelligence, 2003. WI 2003. Proceedings. IEEE/WIC international conference on, pp 372–378

An Efficient Dual Text Steganographic Approach: Hiding Data in a List of Words

Monika Agarwal

Abstract In this paper, we present a novel approach in text steganography using a list of words to hide the secret message. The approach uses the ASCII value of characters and conceals secret message without altering the cover file. The hash value of a message is computed and appended at the end of the message. To provide security, both the message and the appended hash value are encrypted using the proposed encipher algorithm and then the resulting cipher text is embedded inside a cover file using the proposed hide algorithm. After embedding, the stego file, which consists of a list of words, is sent to the receiver. At the receiver side, message authentication and integrity can be verified by applying the same hash function to the deciphered message. The security of the approach is equivalent to the most secure one time pad cryptosystem. We also present an empirical comparison of the proposed approach with some of the existing approaches and show that our approach outperforms the existing approaches.

Keywords Cryptography · Information hiding · Steganography · Text steganography

1 Introduction

Steganography is the art and science of hiding a message inside another message without drawing any suspicion to others so that the message can only be detected by its intended recipient [1]. It is not enough to simply encipher the traffic, as criminals detect, and react to, the presence of encrypted communications [2]. But when steganography is used, even if an eavesdropper gets the stego object,

M. Agarwal (✉)
Department of Computer Science and Engineering,
PDPM-Indian Institute of Information Technology Design and Manufacturing,
Jabalpur, India
e-mail: peace1287@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_47,

he cannot suspect the communication since it is carried out in a concealed way. Steganography, is derived from a finding by Johannes Trithemus (1462–1516) entitled “Steganographia” and comes from the Greek (στεγανό-ς, γραφ − ειν) defined as “covered writing” [3–5]. Steganography gained importance because the US and British government banned the use of cryptography including international mailing of chess games, encrypted mails, clippings from newspapers, etc [6].

Modern steganography is generally understood to deal with electronic media rather than the physical objects and texts [7]. In steganography, the text which we want to conceal is called embedded data, and the text, image, audio, or video file which is used as a medium to hide the text is called cover. The key used (optional) is called stego-key. The resulting object after hiding the data in the cover medium is called stego-object which is transmitted along with the stego-key to its intended recipient. In text steganography, character based text is used to conceal the secret information [8]. Storing text files require less memory and its easier communication makes it preferable to other types of steganographic methods [9].

This paper presents a new approach in text steganography by hiding a message in a list of words. This method works on the ASCII value of a character rather than bits. Hash value of the message is computed and appended at the end of the message and then the message is encrypted using the proposed algorithm. Then for hiding each letter of the encrypted message, a word of certain length is used. At the receiver side, the steps can be performed in reverse order to get back the original message.

The rest of the paper is organized as follows: Sect. 2 describes some of the existing approaches of text steganography. In Sect. 3, the proposed approach is described. Section 4 shows the results of comparison of the proposed approach with the existing approaches. In Sect. 5, we discuss the merits and demerits of the proposed approach and other related issues. Section 6 draws the conclusions.

2 Existing Approaches

In this section, we present some of the popular approaches of text steganography.

2.1 Line Shift

In this method, the secret message is hidden by vertically shifting the text lines to some degree [10, 11]. A line marked has two unmarked control lines one on either side of it for detecting the direction of movement of the marked line [12]. To hide the bit 0, the line is shifted up and to hide the bit 1, the line is shifted down [13].

2.2 Word Shift

In this method, the secret message is hidden by shifting the words horizontally, i.e. left or right to represent the bit 0 or 1 respectively [13].

2.3 Syntactic Method

This technique uses punctuation marks such as full stop (.), comma (,), etc. at proper places to hide the bits 0 and 1 [10, 11, 14].

2.4 White Steg

This technique uses white spaces for hiding the secret message. For example, one space after a word represents the bit 0 and two spaces after a word represents the bit 1 [3, 5, 14].

2.5 Spam Texts

In XML and HTML files, the bit 0 is represented by a lack of space in the tag and the bit 1 is represented by inserting a space inside the tag [13].

2.6 SMS-Texting

In SMS-Texting, to hide the bit 0, full form of the word is used and the bit 1 is hidden by using the abbreviated form of that word [15].

2.7 Feature Coding

In feature coding, the secret message is hidden by altering one or more features of the text [13]. OCR program or re-typing can destroy the hidden information [6, 16].

2.8 SSCE (Secret Steganography Code for Embedding)

This technique first encrypts the secret message using SSCE table and then embed the resulting cipher text in a cover file by inserting articles a or an with the non specific nouns in English language using a certain mapping technique [8].

2.9 Word Mapping Method

This technique first encrypts the secret message using genetic operator crossover and then embed the resulting cipher text in a cover file by inserting blank spaces between words of even or odd length using a certain mapping technique [9].

2.10 MS Word Documents

This technique degenerates the text segments of a document and the secret message is embedded in the choice of degenerations which are then revised with the changes being tracked [17].

2.11 Cricket Match Scorecard

In this method, data is hidden in a cricket match scorecard by pre-appending a meaningless zero before a number to represent bit 1 and leaving the number as it is to represent bit 0 [18].

2.12 CSS (Cascading Style Sheet)

This technique first encrypts the data using RSA public key cryptosystem and then the resulting cipher text is embedded through a Cascading Style Sheet (CSS). A space after a semicolon embeds bit 0 and a tab after a semicolon embeds bit 1 [19].

3 The Proposed Approach

The proposed method hides a message in a list of words. Each letter is hidden in a word of certain length. First, a hash value of the message is computed and appended at the end of the message. The resulting file is then encrypted using a one-time secret key of nearly true random numbers. Then for hiding each character of the cipher text, a word of certain length is taken. The starting letter of a word is determined by masking the sum of digits of the ASCII value of the character to be hidden to an English alphabet. If the sum of digits is 1, then starting letter will be 'a'; if it is 2, then 'b' and so on. At the receiver side, the steps are performed in reverse order to get back the original message. Message authentication and integrity can be verified by calculating hash value of the deciphered message using the same hash function and matching its value against the appended hash value in the deciphered message. Figure 1 shows the proposed model of text steganography. The proposed model has five building blocks: Hash function, which generates a hash code of the message, Encipher function, which scrambles the message using a one-time secret key, Hide function, which hides the encrypted message using a stego key, Seek function, which extracts the hidden information from the stego file using the same stego key, and, Decipher function, which deciphers the extracted message using the same one-time secret key. In all algorithms, we have assumed integer division.

3.1 The Hash Algorithm

1. Get the binary value (a) of the first character of the message.
2. Initial hash value, h = CLS (a), where CLS is circular left shift by one bit.
3. Read next character of the message and convert it to its binary equivalent (b).
4. r = XOR(h, b), where XOR is exclusive OR operation.
5. The next hash value, h = CLS(r).
6. Repeat steps 3–5 till the end of the message.

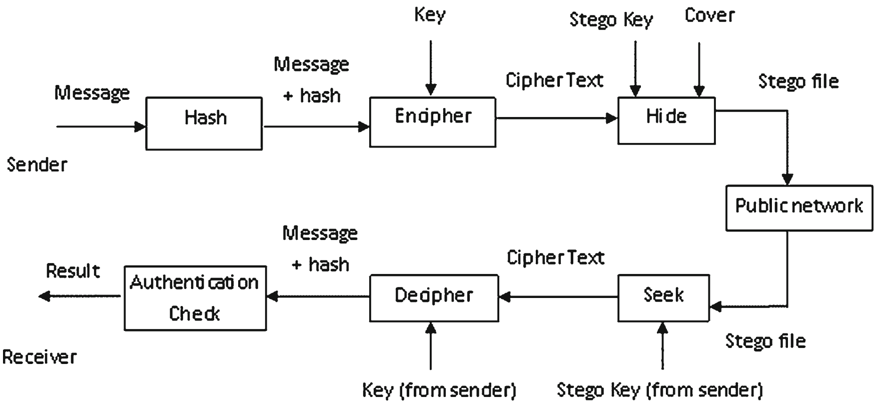


Fig. 1 The proposed model of text steganography

7. Set the most significant bit of h to 0 and then convert it to its decimal representation (n).
8. If n $<$ 30, then n = n + 100.
9. Convert n to its character equivalent and append it at the end of the message.

3.2 The Encipher Algorithm

1. Fill an array A, of size 1000, with random numbers in the range 0–255.
2. Scan a character from the input file and get its ASCII value (n).
3. Generate a random index (i) to the array A and access the value at that index, r = A[i]. Write r in the secret key file.
4. Replace the value (A[i]) with a new random value in the same range.
5. Get s as the sum of square of digits of random number, r.
6. Compute x and y as follows,

$$\mathrm{x} = \mathrm{s}/10.$$

$$\mathrm{y} = \mathrm{s}(\mathrm{mod}10).$$

7. The scrambled value (e),

$$\mathrm{e} = \mathrm{n} - (\mathrm{x} * \mathrm{y}) + \mathrm{r}.$$

8. Write character equivalent of e in cipher text file.
9. Execute steps 2–8 repeatedly till the end of the input file.
10. The secret key is sent to the receiver.

3.3 The Hide Algorithm

1. Calculate length (le) of the input file.
2. Read a letter from the cipher text and get its ASCII value (n).
3. If n $<$ 100, then prefix it with zeroes to make it a three digit number.
4. Calculate k as the most significant digit of n and write it in the stego key file.
5. Length of the word, l = the middle digit of n.
6. If l <6, then l = l + 10.
7. Compute s as the sum of digits of n.
8. Take a word of length l starting from the sth letter in the English alphabet and write it in the stego file.
9. Repeat steps 2 to 8 till the end of the cipher text file.
10. If the length (le) of the cipher text <10, insert 10-le ten letter words in the stego file.

3.4 The Seek Algorithm

1. Read a value (k) from the stego key file.
2. Read a word from the stego file and get its length (l).
3. If $l > 9$, then $l = l - 10$.
4. Compute s by decoding the first letter of the word from the English alphabet.
5. $r = s - (l + k)$.
6. The ASCII value,

$$n = (k^{*}100) + (l^{*}10) + r.$$

7. Convert n to its character equivalent.
8. Repeat above steps till the end of the stego key file.

3.5 The Decipher Algorithm

1. Scan a character from the cipher text file and get its ASCII value, e.
2. Get a value from the secret key (r).
3. Get s as the sum of square of digits of r.
4. Compute x and y as follows,

$$x = s/10.$$

$$y = s(\mathrm{mod}10).$$

5. ASCII value of the character, n, is calculated as,

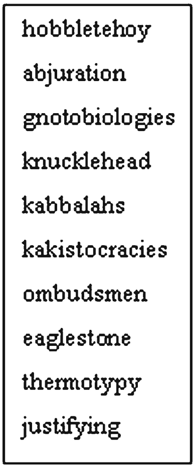
$$n = e + (x^{*}y) - r.$$

6. Convert n to its character equivalent.
7. Execute above steps repeatedly till the end of the cipher text file.

3.6 An Example

Consider the message “pascal” to be hidden in a cover file. The cipher text of the message and its appended hash value comes out to be “ídŔÚĢ H”. Figure 2 shows the stego file after hiding the cipher text.

Fig. 2 The stego file



hobbletehoy
abjuration
gnotobiologies
knucklehead
kabbalahs
kakistocracies
ombudsmen
eaglestone
thermotypy
justifying

4 Experimental Analysis and Results

In this section, we present an experimental comparison of the various text steganographic approaches based on capacity ratio. The capacity ratio is computed by dividing the amount of hidden bytes over the size of the cover text in bytes [20].

Capacity ratio = (amount of hidden bytes)/(size of the cover text media in bytes).

Assuming that one character takes one byte in memory, we have calculated the percentage capacity which is capacity ratio multiplied by 100.
The samples of embedded data used are:

1. Ego (3 byte)
2. Minute (6 byte)
3. Hello World! (12 byte)
4. Failure is never final ! (24 byte)
5. Smile is an inexpensive way to improve your looks. (50 byte)
6. Its not the load that breaks you down, its the way you carry it. (63 byte)
7. Don’t find hundred reasons why you can’t do a thing, but just find one reason why you can and do it. (100 byte)
8. Tide recedes and leaves behind bright sea shells on sand
   Sun sets but its warmth lingers on land
   Music stops and its echoes on in sweet refrains
   For every joy that passes, something beautiful remains (202 byte)
9. Steganography is not a new area. It dates back to 5th century BC. Harpagus used hare to send his message by killing it and hiding the message inside its belly. A person disguised as hunter carried the hare to the destination. Another incident

was of King Darius of Susa. Histiaeus was assigned the duty of shaving the head of his most trusted slave (349 byte).

10. Steganography is not a new area. It dates back to 5th century BC. Harpagus used hare to send his message by killing it and hiding the message inside its belly. A person disguised as hunter carried the hare to the destination. Another incident was of King Darius of Susa. Histiaeus, prisoner of Darius, was assigned the duty of shaving the head of his most trusted slave and then the message was tattooed on his shaved scalp. After some time, when the hairs of the slave grew back, his head was shaved again (508 byte).

Table 1 shows the percentage capacity obtained when our method is applied to the ten experimental samples. We see that for messages having size less than 10 bytes, percentage capacity is low. After 10 bytes, percentage capacity increases and becomes somewhat constant. The rationale behind it is that, in the proposed approach, the stego file consists of at least 10 words regardless of the size of the message. One word can hide one character. If we want to hide the message “eng”, which is of three bytes, three words are required. But, a stego file of three words will raise suspicion. Therefore, minimum size of the stego file is 10 words and hence the percentage capacity is low for messages having size less than 10 bytes. Also, the stego file depends on the ASCII value of characters to be hidden. Since we are concerned with concealing text messages with ASCII value in the range 30–129 and the range of random number is 0–255, therefore, the value of encrypted character falls in the range 1–380. Table 2 shows the average percentage capacity of the approaches when applied to the experimental samples. We observe that the average percentage capacity of our method is greater than the other approaches.

5 Discussion

There are three main issues to be considered when studying steganographic systems: capacity, security, and robustness. Capacity refers to the ability of cover media to store secret data. Security refers to the ability of an eavesdropper to suspect hidden

Table 1 Percentage capacity of the proposed approach over the ten experimental samples

I	II	III	IV	V	VI	VII	VIII	IX	X
3.8	6.7	9.0	9.1	8.9	8.3	8.6	8.8	9.0	8.8

Table 2 Comparison of the av. percentage capacity (>1) of approaches

List of words	White steg	SMS texting	Feature coding	Word mapping	Spam text	Word shift
8.162	1.874	1.71	1.479	1.464	1.164	1.03

information easily. Robustness refers to the ability of protecting the unseen data from modification or corruption [20].

As evident from the experimental results, the average percentage capacity of the proposed approach is much higher than the existing approaches. The stego file does not contain any unnecessary white space. So, if the file is opened with a word processor program, it does not have the possibility of raising suspicion regarding the presence of the concealed information. In case of very small messages (size less than ten bytes), a stego file of three or four words will raise suspicion. Therefore, at least ten words are there in a stego file regardless of the message size. To make it even more secure, it is used in conjunction with cryptography by encrypting a message before concealing it.

Generally, random numbers used by an algorithm are pseudo random numbers. For cryptographic purposes, we need more secure random numbers. True random numbers depend on the environmental conditions such as radioactive decay or CPU temperature, etc. and hence cannot be generated by any hardware or software. So, the encipher algorithm generates a key comprising of nearly true random numbers. Further, the length of the secret key equals the length of the message and the key is discarded after each encryption and decryption. For scrambling the same message again, a different key will be used and thereby resulting in a different cipher text for the same plain text. Thus, if an eavesdropper gets the cipher text or stego file, there is no fixed pattern which he can analyze to figure the hidden content. Hence, the security of the proposed encipher algorithm is equivalent to the one-time pad cryptosystem.

Since no font changing techniques have been applied, the stego file can withstand OCR techniques and retyping does not lead to the loss of the concealed content. Also, the words in a stego file are meaningful and there are no special characters, no extra spaces, and no misspellings. Therefore, the proposed approach is robust. A Limitation of the proposed approach is that it can successfully hide the secret message consisting of characters having ASCII value in the range 30–129, i.e., English alphabets, numbers, and some special characters but we have not considered the entire Unicode list.

6 Conclusion

Steganography although conceals the existence of a message but is not perfectly secure. It is not meant to supersede cryptography but to supplement it. This paper presents a novel approach in text steganography where each character of the embedded data is hidden in a word of cover file without involving any degradation of the cover file. The observed average percentage capacity of the proposed approach is found to be much higher than the existing approaches. The reason behind it is that the approach works on the ASCII value of characters rather than their binary value. The approach is secure because the stego file does not contain any special characters or white spaces and there are no misspelled words. Therefore, opening the stego file with a word processor program will not draw suspicion regarding the existence of concealed content. As no change is done to the cover file while embedding, the cover and stego file are exactly the same. The approach allows further security by scrambling a message, using the one-time pad scheme, before concealing it and thereby leading the message security equivalent to the most secure one-time pad cryptosystem. Apart from this, message integrity can be verified by the use of the hash function. As no font changing techniques have been used, the stego file can withstand OCR techniques and retyping does not lead to the loss of the hidden information. Hence, the proposed approach is robust. The proposed approach can be used to securely transmit sensitive information like PIN, user passwords.

References

1. Changder S, Ghosh D, Debnath NC (2010) Linguistic approach for text steganography through Indian text. In: 2010 2nd international conference on computer technology and development, pp 318–322
2. Anderson RJ, Petitcolas FAP (1998) On the limits of steganography. IEEE J Sel Areas Commun 16(4):474–481
3. Por LY, Delina B (2008) Information hiding—a new approach in text steganography. In: 7th WSEAS international conference on applied computer and applied computational science. Hangzhou China, pp 689–695
4. Petitcolas FAP, Anderson RJ, Kuhn MG (1999) Information hiding—a survey. In: Proc IEEE 87(7):1062–1078
5. Por LY, Ang TF, Delina B (2008) WhiteSteg-a new scheme in information hiding using text steganography. WSEAS Trans Comput 7(6):735–745
6. Rabah K (2004) Steganography-the art of hiding data. Inf Technol J 3(3):245–269
7. Benett K (2004) Linguistic steganography-survey, analysis and robustness concerns for hiding information in text. CERIAS technical report 2004-13
8. Banerjee I, Bhattacharyya S, Sanyal G (2011) Novel text steganography through special code generation. In: International conference on systemics, cybernetics and informatics, pp 298–303
9. Bhattacharyya S, Banerjee I, Sanyal G (2010) A novel approach of secure text based steganography model using word mapping method. Int J Comput Inf Eng 4(2):96–103
10. Shahreza MHS, Shahreza MS (2008) A new dynonym text steganography. In: International conference on intelligent information hiding and multimedia signal processing, pp 1524–1526
11. Shahreza MHS, Shahreza MS (2006) A new approach to persian/arabic text steganography. In: 5th IEEE/ACIS international conference on computer and information science and 1st IEEE/ACIS international workshop on component-based software engineering, software architecture and reuse, pp 310–315
12. Brassil JT, Low SH, Maxemchuk NF, O'Gorman L (1995) Document marking and identification using both line and word shifting. In: Proceedings of INFOCOM '95 proceedings of the fourteenth annual joint conference of the IEEE computer and communication societies, pp 853–860
13. Cummins J, Diskin P, Lau S, Parlett R (2004) Steganography and digital watermarking. School of Computer Science, pp 1–24
14. Bender W, Gruhl D, Morimoto N, Lu A (1996) Techniques for data hiding. IBM Syst J 3(3&4):313–336
15. Shahreza MS, Shahreza MHS (2007) Text steganography in SMS. In: 2007 international conference on convergence information technology, pp 2260–2265

16. Brassil JT, Low S, Maxemchuk NF, O'Gorman L (1995) Electronic marking and identification techniques to discourage document copying. IEEE J Sel Areas Commun 1(8):1495–1504
17. Liu T-Y, Tsai W-H (2007) A new steganographic method for data hiding in microsoft word documents by a change tracking technique. IEEE Trans Inf Forensics Secur 2(1):24–30
18. Khairullah M (2011) A novel text steganography system in cricket match scorecard. Int J Comput Appl 21(9):43–47
19. Kabetta H, Dwiandiyanta BY (2011) Suyoto: information hiding in CSS: a secure scheme text steganography using public key cryptosystem. Int J Cryptogr Inf Secur 1(1):13–22
20. Haidari FA, Gutub A, Kahsah KA, Hamodi J (2009) Improving security and capacity for arabic text steganography using "Kashida" extensions. In: 2009 IEEE/ACS international conference on computer systems and applications, pp 396–399

Secure Cryptosystem with Blind Authentication

Suhas J. Lawand and Madhumita Chatterjee

Abstract Biometric authentication systems are primarily centered on template security, revocability, and privacy. The use of cryptographic primitives to enhance the authentication process addressed some of these concerns shown by biometric cryptosystems. The most common computer authentication method is to use alphanumerical usernames and passwords. This method has significant drawbacks. For example, users tend to pick passwords that can be easily guessed. On the other hand, if a password is hard to guess, then it is often hard to remember. In this paper work, we propose a provably secure and blind biometric authentication protocol, which addresses the concerns of user's privacy, template protection, and trust issues.

Keywords Blind · RSA · Template · Privacy

1 Introduction

Humans recognize each other according to their various characteristics for ages. We recognize others by their face when we meet them and by their voice as we speak to them. Identity verification (authentication) in computer systems has been traditionally based on something that one has (key, magnetic or chip card) or one knows (PIN, password) [1, 2]. Things like keys or cards, however, tend to get stolen or lost and passwords are often forgotten or disclosed. To achieve more reliable verification or identification we should use something that really characterizes the given person. Biometrics offer automated methods of identity verification or identification

S. J. Lawand (✉) · M. Chatterjee
Department of Computer Engineering, Pillai's Institute of Information Technology,
Navi Mumbai, India
e-mail: suhas_lawand@yahoo.com

M. Chatterjee (✉)
e-mail: c_a_mita@yahoo.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_48,

on the principle of measurable physiological or behavioral characteristics such as a fingerprint or a voice sample. The characteristics are measurable and unique. Biometrics is the study of automated methods for uniquely recognizing humans based upon one or more intrinsic physical or behavioral traits [3, 4]. In other words, the study of Biometrics explores ways to distinguish between individuals using physical characteristics (things we are) and personal traits (things we do). The most common physical characteristics explored and used are facial features, eyes (iris and retina), fingerprints and hand geometry. Handwriting and voice are examples of personal traits which could be used to distinguish between individuals. The described characteristics and traits can be used to identify different individuals, because they all satisfy specific requirements. They are all universal and unique, which means that everybody has them and that the characteristics or traits are different for any two individuals. In addition to that they are all more or less permanently. Fingerprints and face recognition are the two most common used characteristics to distinguish between individuals [3].

To address the problem of the ease of guessing passwords, we proposed a system where images are used as a password. The protocol is blind in the sense that it reveals only the identity, and no additional information about the user or the biometric to the authenticating server or vice-versa. As the protocol is based on asymmetric encryption of the biometric data, it captures the advantages of biometric authentication as well as the security of public key cryptography. It is double blind authentication as user creates a password by clicking on sequence of images and proceeds for biometric authentication.

2 Literature Survey

Human factors are often considered the weakest link in a computer security system. There are three major areas where human-computer interaction is important: authentication, security operations, and developing secure systems. The most common computer authentication method is for a user to submit a user name and a text password. The vulnerabilities of this method have been well known. One of the main problems is the difficulty of remembering passwords. Studies have shown that users tend to pick short passwords or passwords that are easy to remember [5]. Unfortunately, these passwords can also be easily guessed or broken.

Graphical password schemes have been proposed as a possible alternative to text-based schemes, motivated partially by the fact that humans can remember pictures better than text. Pictures are generally easier to be remembered or recognized than text. In addition, if the number of possible pictures is sufficiently large, the possible password space of a graphical password scheme may exceed that of text based schemes and thus presumably offer better resistance to dictionary attacks. Because of these (presumed) advantages, there is a growing interest in graphical password.

The encryption-based security of biometric templates tends to model the problem as that of building a classification system that separates samples in the encrypted domain [1, 6, 7]. However, a strong encryption mechanism destroys any pattern in the data, which adversely affects the accuracy of verification. Hence, any such matching mechanism necessarily makes a compromise between template security (strong encryption) and accuracy (retaining patterns in the data). The template protection can be achieved by four categories of solutions [8]. The first category is salting, which offers security using a transformation function seeded by a user specific key.

The second category is noninvertible transform which applies a trait specific noninvertible function on the biometric template so as to secure it. The parameters of the transformation function are defined by a key which must be available at the time of authentication [9, 10]. However, these methods are often biometric specific and do not make any guarantees on preservation of privacy, especially when the server is not trusted.

The third and fourth category is based on biometric cryptosystem, which try to integrate the advantages of both biometric and cryptography to enhance the overall security and privacy of an authentication system. Such systems used the biometric as a protection for a secret key (key binding approach) or use the biometric data to directly generate a secret key (key generation approach). However, this would become a key-based authentication scheme and would lose the primary advantage of biometric authentication, which is its non-reputable nature.

3 Problem Statement

Biometric authentication systems are gaining wide-spread popularity in recent years due to the advances in sensor technologies as well as improvements in the matching algorithms that make the systems both secure and cost-effective. They are ideally suited for both high security and remote authentication applications due to the non-reputable nature and user convenience. Most biometric systems assume that the template in the system is secure due to human supervision (e.g., immigration checks and criminal database search) or physical protection (e.g., laptop locks and door locks). However, a variety of applications of authentication need to work over a partially secure or insecure network such as ATM networks or the Internet. Authentication over insecure public networks or with untrusted servers raises more concerns in privacy and security. The primary concern is related to the security of the plain biometric templates, which cannot be replaced, once they are compromised. The privacy concerns arise from the fact that the biometric samples reveal more information about its owner (medical, food habits, etc.) in addition to the identity [1].

In this paper, we propose a system to provide two factor authentications to enhance security of biometric data. The first factor is graphical password scheme where password is a set of images. Graphical password schemes is a possible alternative to text-based schemes [5, 11] motivated by the fact that humans can remember pictures better than text studies also supports the fact that pictures are generally easier to be remembered or recognized than text. Second factor is biometric authentication where client and server do not reveal any information to each other.

4 Proposed System

Biometrics is the science of verifying and establishing the identity of an individual through physiological features or behavioral traits. Security of biometric data is essential as biometric data does not change over course of life. The objective of proposed authentication system is to provide security to biometric data through graphical password scheme.

4.1 Double Blind Authentication System

Blind Authentication is a biometric authentication protocol that does not reveal any information about the biometric samples to the authenticating server. It also does not reveal any information regarding the server, to the user or client. User creates a graphical password by clicking on sequence of images as shown in Fig. 1. After creating the graphical password, user proceeds for biometric registration. When a user login with correct graphical password authentication of biometric is done as shown in Fig. 2.

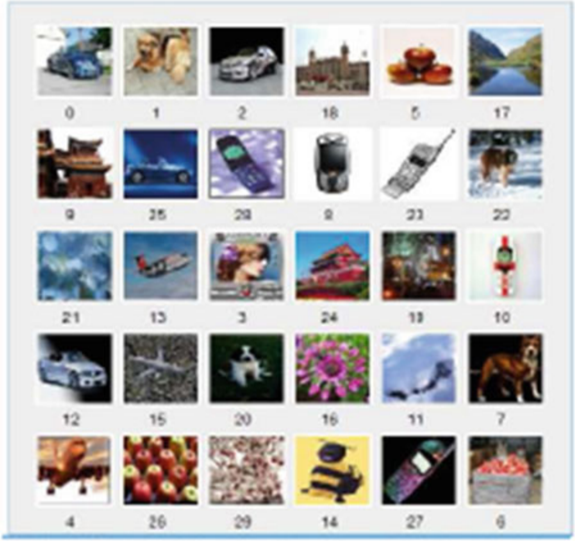


Fig. 1 Graphical password authentication method

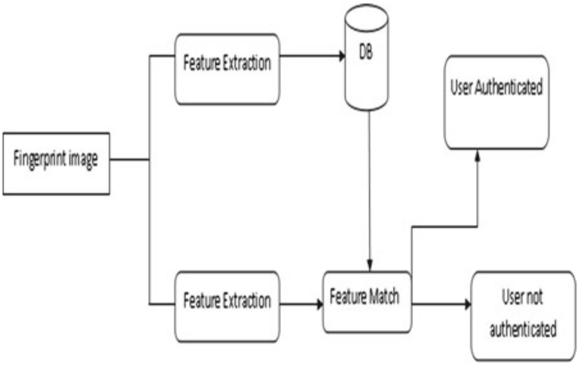


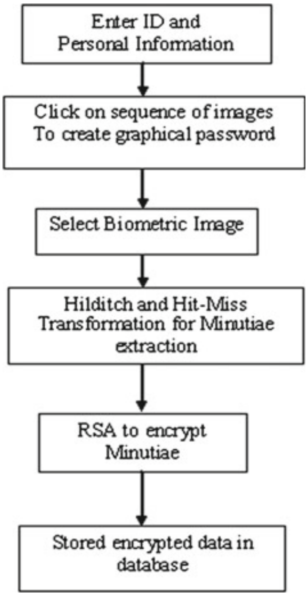
Fig. 2 Biometric authentication

4.2 Working of Proposed System

The system has two factor authentications. First factor is graphical password scheme and second is biometric authentication. In graphical password scheme there are two phases creating phase and login phase. In create phase a graphical password is created by clicking on sequence of images. If user successfully create graphical password, biometric registration process starts. In biometric registration process features are extracted from image by applying Hit-miss algorithm and Hilditch algorithm to extract correct minutiae points. Each sub iteration begin by examining the neighborhood of each pixel in binary image and based on pixel selection criteria it checks whether pixel can deleted or not. The sub iteration continues until no more pixels can be deleted. Minutiae points are those having pixel value one (ridge ending) [12, 13] as their neighbor or more than two ones (ridge bifurcations) in their neighborhood. These minutiae points are encrypted using RSA algorithm and stored at server side. In login phase if users able to follow same sequence of images as that of create phase, they are able to login and able to proceed for biometric authentication process. In authentication of biometric features are extracted from finger print image and compare with register image, if match is found user is authenticated and new application is open. If features are not match user is not authenticated.

The graphical password provides security from shoulder surfing attack. The sequence of images user click for creating graphical password is shuffle during login phase. To shuffle images Fisher Yates algorithm is used. If user creates a graphical password by following particular sequence of images during sign up phase, he can sign in by following same sequence but in between user can click on any other images which provides security from shoulder surfing. If user forgets the graphical password, user can reset his password. For resetting password user has to enter Id, a secret code is created and send to register mobile number. By entering the secret code user can reset graphical password.

Fig. 3 Flowchart of registration process


Enter ID and Personal Information
Click on sequence of images To create graphical password
Select Biometric Image
Hilditch and Hit-Miss Transformation for Minutiae extraction
RSA to encrypt Minutiae
Stored encrypted data in database

4.3 *Proposed Algorithm*

We propose a system where user has to create a graphical password by clicking on sequence of images as shown in Fig. 3. Once user successfully create graphical password he can proceed for biometric registration.

In authentication Phase (see Fig. 4) the user has to follow same sequence of images to authenticate. If user successfully pass graphical phase he can proceed for biometric authentication. In order to authenticate as authorized user, user has to pass through two levels of authentication.

5 Result

The Proposed system provides double blind authentication. It starts with graphical password authentication technique. In graphical password authentication, the specified number of images is randomly selected by the system from a database to form an image portfolio as shown. User has to click on images to generate graphical password as shown in Fig. 5.

Once user successfully create graphical password, user can proceed for biometric authentication as shown in Fig. 6

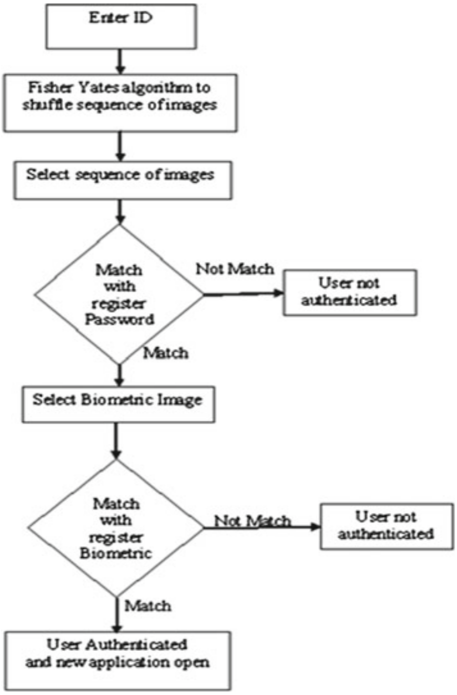

Enter ID
Fisher Yates algorithm to shuffle sequence of images
Select sequence of images
Match with register Password
Not Match
User not authenticated
Match
Select Biometric Image
Match with register Biometric
Not Match
User not authenticated
Match
User Authenticated and new application open

Fig. 4 Flowchart of authentication process

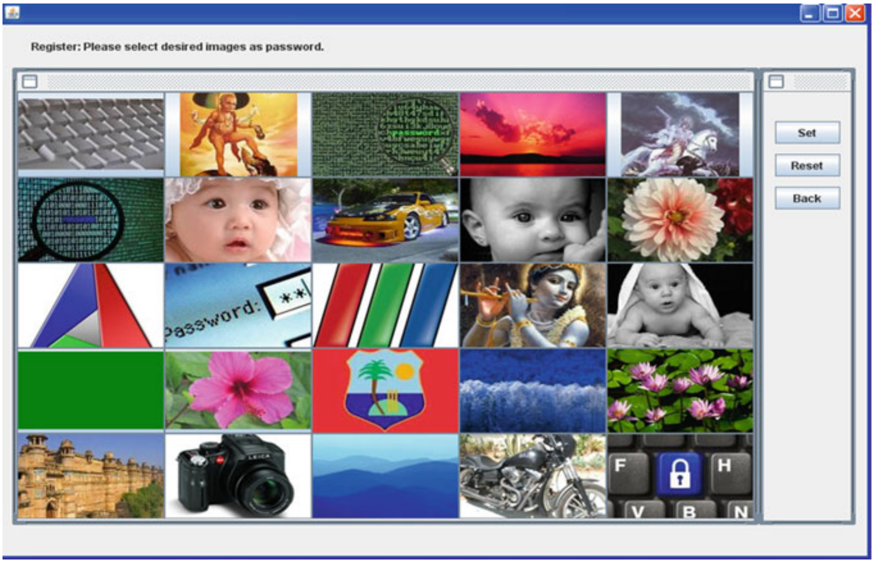
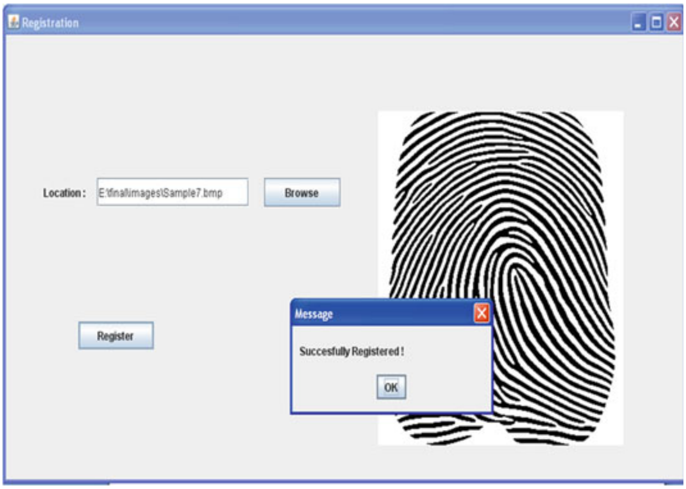

Register: Please select desired images as password.

Fig. 5 Images to create graphical password



Location : E:\final\images\Sample7.bmp

Succesfully Registered !

Fig. 6 Biometric image registered successfully

5.1 Analysis

The Biometric system although considered as the greatest identification system, still is not without certain loopholes. The fake physical biometric have been designed to circumvent the whole system. The biggest threat is that biometric spoofing can be easily done with little or no technical knowledge at all. Moreover the materials for creating such false biometrics are very cheap and easily obtainable. Another factor is the time period of attack, which can occur at the point of entry to the system. The greatest damage is dealt in the case of biometrics like fingerprint, hand and iris as extensive traces of biometric can be easily obtain from objects used in daily life making original biometric easily available with or without owner of biometric.

5.2 Resilience of the System to Attacks

The proposed system was tested on various conditions are as follows

Test Case 1: Unauthorized user hack graphical password. In such scenario unauthorized user able to process through graphical authentication but has to give his fingerprint image for biometric authentication. The biometric sample of unauthorized user doesn't match with registered image and authentication failure occurs.

Test Case 2: Attacker with fake biometric If attacker obtain fake biometric of authorized user, attacker has to process through graphical password authentication.

If attacker unable to process though graphical password phase then the fake biometric of authorized user is of no use.

Test Case 3: Authorized user forgets graphical password. In such scenario user is given a chance to set graphical password again. After resetting the graphical password, user has to process though biometric authentication process and is then duly authenticated.

The above test cases show that with graphical password scheme fake physical biometric attack can be easily ruled out. The graphical password scheme design to protect biometric data is free from shoulder surfing attack which enhances security of system.

6 Conclusion

User authentication is a fundamental component in most computer security. In this paper work, we proposed a graphical password plus biometric authentication system. The system combines graphical password and biometric data for authentication. It also provides multi-factor authentication in a friendly intuitive system. The primary advantage of the system is that the authentication server need not know the specific biometric trait that is used by a particular user, which can even vary across users. Once a registration process encrypts the specific biometric of a person, the authentication server is verifying the identity of a user with respect to that encryption. The real identity of the person is hence not revealed to the server, making the protocol, completely blind. This allows one to revoke enrolled templates by changing the encryption key, as well as use multiple keys across different servers to avoid being tracked, thus leading to better privacy.

References

1. Upmanyu M, Namboodiri AM, Srinathan K, Jawahar CV (2004) Blind Authentication. IEEE Trans Inf Forensics Secur 5(2)
2. Ratha N, Chikkerur S, Connell J, Bolle R (2007) Generating cancelable fingerprint templates. IEEE Trans Pattern Anal Mach Intell 29(4):561–572
3. Savvides M, Kumar BV (2004) Cancellable biometric filters for face recognition. Int Conf Pattern Recogn (ICPR) 3:922–925
4. Kong A, Cheung K, Zhang D, Kamel M, You J (2006) An analysis of biohashing and its variants. Pattern Recogn 39(7):1359–1368
5. Biddle R, Chiasson S, van Oorschot PC (2009) Graphical passwords: learning from the first generation version: October 2 Technical report TR-09-09. Carleton University, Ottawa, Canada, School of Computer Science
6. Jain AK, Nandakumar K, Nagar A (2008) Biometric template security. EURASIP 8(2):1–17
7. Upmanyu M, Namboodiri AM, Srinathan K, Jawahar CV (2009) Efficient biometric verification in the encrypted domain. In: 3rd International conference biometrics, pp 906–915

8. Uludag U, Pankanti S, Prabhakar S, Jain AK (2004) Biometriccryptosystems: issues and challenges. Proc IEEE 92(6):948–960
9. Teoh A, Jin B, Connie T, Ngo D, Ling C (2006) Remarks on bioHash and its mathematical foundation. Inf Process Lett 100(4):145–150
10. Boult T, Scheirer W, Woodworth R (2007) Revocable fingerprint biotokens: accuracy and security analysis. In: IEEE conference computer vision and pattern recognition (CVPR), pp 1–8
11. Suo X, Zhu Y, Scott G Graphical passwords: A survey. Owen Department of Computer Science Georgia State University xsuo@student.gsu.edu, yzhu@cs.gsu.edu, owen@siggraph.org
12. Bansal R, Sehgal P, Bedi P (2010) Effective morphological extraction of true fingerprint minutiae based on the hit or miss transform. Int J Biometrics Bioinform (IJBB) 4(2):71–85
13. Kaur M, Singh M, Girdhar A, Sandhu PS (2008) Fingerprint verification system using minutiae extraction technique. World Acad Sci Eng Technol 46:497–502

An Effective Technique for Intrusion Detection Using Neuro-Fuzzy and Radial SVM Classifier

A. M. Chandrasekhar and K. Raghuveer

Abstract Intrusion detection is not yet a perfect technology. This has given data mining the opportunity to make several important contributions to the field of intrusion detection. In this paper, we have proposed a new technique by utilizing data mining techniques such as neuro-fuzzy and radial basis support vector machine (SVM) for the intrusion detection system. The proposed technique has four major steps in which, first step is to perform the Fuzzy C-means clustering (FCM). Then, neuro-fuzzy is trained, such that each of the data point is trained with the corresponding neuro-fuzzy classifier associated with the cluster. Subsequently, a vector for SVM classification is formed and in the fourth step, classification using radial SVM is performed to detect intrusion has happened or not. Data set used is the KDD cup 99 dataset and we have used sensitivity, specificity and accuracy as the evaluation metrics parameters. Our technique could achieve better accuracy for all types of intrusions. It achieved about 98.94 % accuracy in case of DOS attack and reached heights of 97.11 % accuracy in case of PROBE attack. In case of R2L and U2R attacks it has attained 97.78 and 97.80 % accuracy respectively. We compared the proposed technique with the other existing state of art techniques. These comparisons proved the effectiveness of our technique.

Keywords Intrusion detection system · Neuro-fuzzy · Radial support vector machine · Fuzzy C-Means clustering

A. M. Chandrasekhar (✉)
Department of Computer Science,
Sri Jayachamarajendra college of Engineering (SJCE),
Mysore, Karnataka 570006, India
e-mail: amblechandru@gmail.com

K. Raghuveer
Department of Information Science, National Institute of Engineering (NIE),
Mysore, Karnataka 570008, India
e-mail: raghunie@yahoo.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_49,

1 Introduction

Intrusion Detection Systems (IDS) form a key part of system defense, where it identifies abnormal activities happening in a computer system. Intrusion detection system has been an active area of research and development for the past few decades. This is primarily because of the mounting of attacks on computers and on networks in recent years and computerized scrutiny has become a compulsory addition to IT security [1]. A typical intrusion detection system consists of three functional components: an information source that provides a stream of event records, an analysis Engine that determines signs of intrusions and a decision maker that applies a number of rules on the outcomes of the analysis engine, and resolves what reactions should be done based on the outcomes of the analysis engine [2]. There are three categories of intrusion detection systems which are host based where information is found on a single or multiple host systems, network based that examines the information captured from network communications and vulnerability assessment based that identifies vulnerabilities in internal networks and firewall. Whereas based on the functionality, intrusion detection can be classified into two types: anomaly detection and misuse detection [3]. Based on a study of latest research literatures, there are quite a lot of research that attempts to relate data mining and machine learning techniques to the intrusion detection systems so as to design more intelligent intrusion detection model. Currently the support vector learning technique is featuring superior [4].

The rest of the paper is organized as follows: The proposed technique for intrusion detection using neuro-fuzzy and radial SVM classifier is presented in Sect. 2. The detailed experimental setup, data set description, results, comparative study are discussed in Sect. 3. The conclusions are summed up in Sect. 4.

2 Our Proposed Technique

The proposed Intrusion detection technique is majorly a four step methodology, which are namely: (1) Clustering using Fuzzy C-Means (FCM) Clustering, where the input data set is clustered into K-clusters where, K is the number of clusters desired, (2) Neuro-fuzzy training, where each of the data in a particular cluster is trained with the respective neural network associated with each of the cluster, (3) Generation of vector for SVM classification, which consists of attribute values obtained by passing each of the data through all of the trained Neuro-fuzzy classifiers, and an additional attribute which has membership value of each of the data and (4) Classification using radial SVM to detect intrusion has happened or not. The block diagram of the proposed technique is given in Fig. 1.

The dataset given as input for intrusion detection technique consists of large number of data, where each of the data considered has numerous attributes associated with it. Hence, to perform classification considering all these attributes is a hectic and time consuming task. Processing and executing this lump amount of data also results in

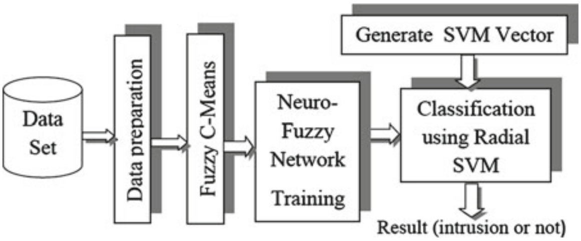

Data Set
Data preparation
Fuzzy C-Means
Neuro-Fuzzy Network Training
Generate SVM Vector
Classification using Radial SVM
Result (intrusion or not)

Fig. 1 Block diagram of proposed technique

increasing the error rate and also negatively affects efficacy of the classifier system. In order to overcome this problem, our proposed technique comes up with a solution where the number of attributes defining each of the data is reduced to a small number through a sequence of steps. This process ultimately results in making the intrusion detection more efficient and also yields a less complex system with a better result. Data set used to evaluate the validity of the proposed technique is prepared from the KDD Cup 99 data set and the detailed explanation of it is given in Sect. 3.

2.1 Clustering Module

Fuzzy C-Means (FCM) algorithm is a technique of clustering which permits one piece of data to two or more clusters. This technique was established by Dunn [5] and is mostly employed in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N} \sum_{j-1}^{C} \mu_{ij}^{m} \|x_j - c_j\|^2 \quad 1 \le m < \infty$$

where m is any real number greater than 1. μ_{ij} is the degree of membership of x_i in the cluster j, x_i the *i*th of dimensional measured data, c_j is the d-dimension centre of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the centre. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership μ_{ij} and the cluster centers by c_j.

In this technique, initially the input data is grouped into clusters by use of FCM algorithm. Our input data set consists of the normal data and different types of intrusions like DOS, PROBE, R2L and U2R. Therefore, clustering results in grouping these data into clusters based on the type of intrusions. Examining and learning the behavior and characteristics of a single data point within a cluster can give hints and clue on all other data points in the same cluster. This is because of the fact that all data points inside a cluster differ only by a small amount and usually follow a more or less similar structure. Hence, clustering the data and then classifying is a simpler method and is less time consuming. We have employed FCM clustering as time incurred is less when compared to hierarchical clustering and yields a better result when compared to K-Means clustering [6]. In our case, the data set is given as an input to the clustering process. After FCM clustering, it capitulate K clusters wherein each cluster will be a type of the intrusion except for the one with normal data.

2.2 Neuro-Fuzzy Classifier Module

Neural networks are a significant tool for classification. But it has many disadvantages of having impossible interpretation of the functionality and also faces difficulty in determining the number of layers and the number of neurons [7]. These disadvantages can be overcome by incorporating fuzzy into neural networks and results in better results and outcomes. Neuro-fuzzy hybridization is widely termed as Fuzzy Neural Network (FNN) or Neuro-Fuzzy System (NFS) in the literature [8]. Neuro-fuzzy refers to the combination of fuzzy set theory and neural networks with the advantages of both. Neuro-fuzzy incorporates fuzzy sets and a linguistic model consisting of a set of IF-THEN fuzzy rules. The main strength of neuro-fuzzy systems is that they are universal approximators with the ability to solicit interpretable IF-THEN rules. The main advantages of using neuro-fuzzy are that it can handle any kind of information (numeric, linguistic, logical, etc.). It can manage imprecise, partial, vague or imperfect information. It can resolve conflicts by collaboration and aggregation. It has self-learning, self-organizing and self-tuning capabilities. There is no need of prior knowledge of relationships of data mimic human decision making process. It can perform fast computation using fuzzy number operations.

Fuzzy C-Means clustering results in the formation of K-clusters where each cluster will be a type of intrusion or the normal data. For every cluster, we have a neuro-fuzzy classifier associated with it. That is, there will be K number of neuro-fuzzy classifiers for K number of clusters formed. Each neuro-fuzzy classifier is trained with the data in the respective cluster. Neuro-fuzzy makes use of back-propagation learning to find out the input membership function parameters and the least mean square method to find out the consequent parameters.

The main purpose of the proposed technique is to decrease the number of attributes associated with each data, so that classification can be made in a simpler and easier way. Neuro-fuzzy classifier is employed to efficiently decrease the number of attributes. Classification of the data point considering all its attributes is a very difficult task and takes much time for the processing, hence decreasing the number of attributes related with each of the data point is of paramount importance.

2.3 SVM Training Vector Module

Support Vector Machine is used here as it achieves enhanced results when contrasted to other classification techniques especially when it comes to binary classification. The input data is trained with neuro-fuzzy after the initial clustering as we have discussed earlier, then the vector necessary for the SVM is generated. The SVM vector array $\mathrm{S} = \{\mathrm{D}_1, \mathrm{D}_2, \ldots \mathrm{D}_\mathrm{N}\}$ where, D_i is the *i*th data and N is a total number of input data. Here, after training through the neuro-fuzzy the attribute number reduces to K numbers. $\mathrm{D}_\mathrm{i} = \{\mathrm{a}_1, \mathrm{a}_2, \ldots \mathrm{a}_\mathrm{k}\}$, Here the D_i is the *i*th data governed by attribute values a_i, where a_i will have the value after passing through the *i*th neuro-fuzzy. Total number of neuro-fuzzy classifiers trained will be K, corresponding to the K clusters formed after clustering. Initially, we have used the FCM clustering which is error prone and does not yield the exact values. Hence in order to overcome this and to have a better result we include a parameter known as membership value. Inclusion of the membership value into the attribute list results in a better performance of the classifier. Membership value μ_{ij} is defined as given by the equation below

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{||x_i - c_i||}{||x_i - c_k||} \right)^{\frac{2}{m-1}}}$$

Hence the SVM vector is modified as $\mathrm{S}^* = \{\mathrm{D}^*{}_1, \mathrm{D}^*{}_2, \ldots \mathrm{D}^*{}_\mathrm{N}\}$ where S^* is the modified SVM vector which consists of modified data $\mathrm{D}^*{}_\mathrm{i}$, which consists of an extra attribute of membership value μ_{ij}. $\mathrm{D}_1{}^* = \{\mathrm{a}_1, \mathrm{a}_2, \ldots \mathrm{a}_\mathrm{k}, \mu_{\mathrm{ij}}\}$, hence the attribute number is reduced to K + 1 where K is the number of clusters. This results in easy processing in the final SVM classification. This is due to the fact that input data which had 34 attributes is now constrained to K + 1 attributes. This also reduces the system complexity and time incurred.

2.4 Radial-SVM Classifier Module

SVM classifier is used as it produces better results for binary classification when compared to the other classifiers. But use of linear SVM has the disadvantages of getting less accuracy result, over fitting results and robust to noise. These short comings are effectively suppressed by the use of the Radial SVM where nonlinear kernel functions are used and the resulting maximum-margin hyper-plane fits in a transformed feature space. Hilbert space of infinite dimensions is formed as the corresponding feature space when the kernel used is a Gaussian radial basis function. In our proposed technique, nonlinear kernel functions are used and the resulting maximum-margin hyper-plane fits in a transformed feature space. When the kernel used is a Gaussian radial basis function, the corresponding feature space is a Hilbert space of infinite dimensions. The Gaussian Radial Basics function is given by the equation

$$k(x_i, x_j) = \exp(-\gamma||x_i - x_j||^2, \quad for \quad \gamma > 0, \gamma = 1/2\sigma^2$$

The input dataset having large number of attributes is changed into data having K + 1 attributes by performing the above steps. The data with constrained number of attributes is given to the radial SVM, which is binary classified to detect if there is intrusion or not.

3 Experimental Setup and Results

The complicated version of DARPA dataset which encircle only network data (TCP dump) is named as KDD Cup 99 dataset. This dataset consists of moderately about 4,900,000 single correlation vectors where each single connection vector consists of 41 features and is marked as either normal or an attack, with accurately one particular attack type [9]. A wide range of attacks integrated in the dataset come beneath the four main categories: (1) Denial of Service Attacks (DOS): This attack is an attack where the attacker creates a few calculations or memory resource completely engaged or out of stock to handle authentic requirements, or reject justifiable users the right to utilize a machine. (2) User to Root Attacks (U2R): In this category of attack the attacker begins by accessing a normal user account on the system and get advantage of several vulnerability to accomplish root access to the system. (3) Remote to Local (R2L) Attacks: This attack takes place when an attacker who has the potential to send packets to a machine over a network but does not have an account on that machine, makes use of some vulnerability to accomplish local access as a user of that machine. (4) Probes/scans (PROBE): Probing is a collection of attacks where an attacker scrutinizes a network to gather information or to conclude prominent vulnerabilities. A complete listing of the set of features given in KDD Cup 99 dataset defined for the connection records and types of attacks falling into four major categories are given in [10].

The proposed technique is implemented using MATLAB. For the testing and evaluation of proposed method we have taken KDD Cup 99 dataset and to rating its performance, we made use of sensitivity, specificity and accuracy as metrics. It is very tough to perform the proposed technique on the KDD cup 99 dataset for estimating the performance, because it is of large scale. The subset of 10 % of KDD Cup 99 dataset is made use for our experimentation. The number of data points taken for training and testing phase is given in Table 1. Totally, in training, we considered 26114 data points and in testing we considered 27112 data points.

In this paper we used confusion matrix: false positive (FP), false negative (FN), true positive (TP), and true negative (TN). The performance of a binary classification test is statistically measured by sensitivity and specificity. The confusion matrix is calculated for both training and testing dataset in the testing phase. The obtained result for all attacks and normal data are given in Table 2. The results are evaluated

Table 1 Dataset and equations used during evaluation

Attack	Training	Testing	Equations used
Normal	12500	12500	$sensitivity = TP/(TP + FN)$
DOS	12500	12500	$specificity = TN/(TN + FP)$
Probe	1054	2053	$Accuracy = \frac{(\mathrm{TN} + \mathrm{TP})}{\mathrm{TN} + \mathrm{TP} + \mathrm{FN} + \mathrm{FP}}$
R2L	39	38	
U2R	21	21	

Table 2 Experimental results obtained for the training and testing dataset

Metrics	Types of attacks							
	DOS		Probe		R2L		U2R	
	Train	Test	Train	Test	Train	Test	Train	Test
TN	12491	12235	12491	12235	12491	12235	12491	12235
FP	9	265	9	265	9	265	9	265
TP	12499	12500	2023	1897	39	25	14	11
FN	1	0	31	156	0	13	7	10
Specificity	99.93	97.89	99.93	97.88	99.93	97.88	99.93	97.88
Sensitivity	99.99	1	98.491	92.401	1	65.79	66.67	52.38
Accuracy	99.96	98.94	99.725	97.11	99.93	97.78	99.87	97.80

with the evaluation metrics namely, sensitivity, specificity and accuracy [11]. So as to discover these metrics, we first compute confusion matrix then we apply values into the equations shown in Table 1 to find sensitivity, specificity and accuracy.

From the table, it is clear that we have achieved about 99 % accuracy in case of DOS attack and reached heights of 97 % accuracy in case of PROBE attack. In the case of R2L and U2R attacks it has attained 97.8 % accuracy.

Table 3 demonstrates the comparison of our technique with other state of art technique. Form the Table 3, it is apparent that the our technique has obtained a reliable peak scores for all types of intrusions. In the case of DOS intrusion, we have attained 98.94 %, which is the maximum accuracy value when compared to other methods. In the case of the PROBE intrusion, we achieve a very good result of 97.11 % accuracy. For both U2R and R2L, once again we have the maximum

Table 3 Accuracy comparison (in percentage) with existing methods

Attack	Different methods							
types	[12]	[13]	[14]	[15]	[16]	[17]	[18]	Proposed
Probe	83.3	73.2	75	98.60	96.7	81.4	99.3	97.11
DOS	97.1	96.9	96.8	97.40	24.3	60.0	98.1	98.94
U2R	13.2	6.6	5.3	86.30	81.8	58.8	89.7	97.80
R2L	8.4	10.7	4.2	29.60	5.9	24.2	48.2	97.78

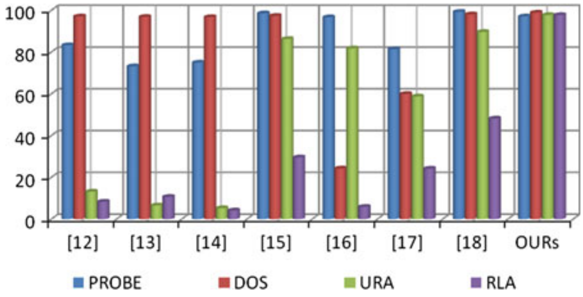


Fig. 2 Plot of the performance of various methods

value of 97.78 and 97.80 % accuracy respectively when compared to others. Our technique which use both the neural network and SVM classifier performs well in all types of intrusions. We have received the best results as we have employed fuzzy-neural networks to diminish the number of attributes of the data and by the utilization of radial SVM in the final classification. Figure 2 shows the comparative results in graphical form.

4 Conclusion

This paper presents an efficient technique for intrusion detection by making use of fuzzy-neural networks and radial support vector machine. The proposed technique consists of initial clustering, fuzzy neural network training, formation of SVM vector and the final classification using the radial SVM. KDD cup 99 dataset is used for experimental verification. Here, we have used confusion matrix for the purpose of evaluation of our proposed technique and the results are evaluated with the evaluation metrics namely, sensitivity, specificity and accuracy. We have received the best results and these are compared with results of other existing methods and from that, it is clear that our proposed technique outperformed all other state of art techniques.

References

1. Dubey GP, Gupta N, Bhujade RK (2011) A novel approach to intrusion detection system using rough set theory and incremental SVM. In: IJSCE, vol 1, No 1, pp 14–18 (2011)
2. Yao JT, Zhao SL, Saxton LV (2005) A study on fuzzy intrusion detection. In: Proceedings of data mining, intrusion detection, information assurance, and data networks security SPIE, vol 5812, pp 23–30 (2005)
3. Wagner D, Soto P (2002) Mimicry attacks on host based intrusion detection systems. In: Proceedings of the 9th ACM conference on computer and communications, security, pp 255–264 (2002)

4. Lee H, Song J, Park D (2005) Intrusion detection system based on multi-class SVM. Dept. of computer & Info Science, Korea Univ., pp 511–519 (2005)
5. Dunn JC (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J Cybern 3:32–57
6. Chandrashekhar AM, Raguveer K (2012) Performance evaluation of data clustering techniques using KDD cup 99 intrusion data set. In: IJINS, vol 1, No 4, pp 294–305 (2012)
7. Vieira J, Morgado Dias F, Mota A (2004) Neuro-fuzzy systems, a survey. In: Proceedings international conferenceon neural networks and applications (2004)
8. Jang R (1992) Neuro-fuzzy modelling: architectures, analysis and applications. Ph.D. thesis, University of California, Berkley (1992)
9. Tavallaee M, Bagheri E, Lu W, Ghorbani AA (2009) A detailed analysis of the KDD CUP 99 data set. In: Proceedings IEEE international conference on computational intelligence for security and defence applications. Ottawa, Ontario, Canada, pp 53–58 (2009)
10. Chebrolu S et al (2005) Feature deduction and ensemble design of intrusion detection systems. In: Elsevier journal of computers and security, vol 24, No 4, pp 295–307 (2005)
11. Zhu W, Zeng N, Wang N (2010) Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS® implementations. In: NESUG proceedings: health care and life sciences, Baltimore, Maryland (2010)
12. Pfahringer B (2000) Winning the KDD99 classification cup bagged boosting. In: SIGKDD explorations, vol 1, pp 65–66 (2000)
13. Agarwal R, Joshi MV (2000) PNrule: a new framework for learning classifier models in data mining. In: Case-study in network intrusion detection (2000)
14. Ambwani T (2003) Multi class support vector machine implementation to intrusion detection. In: Proceedings of IJCNN, pp 2300–2305 (2003)
15. Gupta KK, Nath B, Kotagiri R (2008) Layered approach using conditional random fields for intrusion detection. In: IEEE transactions on dependable and secure computing, vol 5 (2008)
16. Lee W, Stolfo S (2000) A framework for constructing features and models for intrusion detection systems. In: Information and system security, vol 4, pp 227–261 (2000)
17. Lee JH, Lee J-H, Sohn S-G et al (2008) Effective value of decision tree with KDD 99 intrusion detection datasets for intrusion detection system. In: Proceedings of international conference on advanced communication technology, vol 2, pp 1170–1175 (2008)
18. Tran TP, Cao L, Tran D et al (2009) Novel intrusion detection using probabilistic neural network and adaptive boosting. In: IJCSIS, vol 6, No 1, pp 83–91 (2009)

A Novel Octuple Images Encryption Algorithm Using Chaos in Wavelet Domain

Musheer Ahmad, Bashir Alam, Arpit Jain and Vipul Khare

Abstract The advancements in the network and multimedia technologies have made information security more exigent and demanding. It brings new challenges to develop security methods that are credential enough to encrypt a number of images and generate a single encrypted image containing the information of all plain-images. Here, we propose a novel image encryption algorithm which has the efficacy of encrypting eight distinct plain-images simultaneously. Low frequency components of plain-images are selected and processed. The algorithm makes use of three chaotic systems to get visual effect of disorder, distorted and indistinguishable single encrypted image. The encryption/decryption processing operations such as chaos-based circular rotation and random mixing of pixels are carried out in wavelet domains. The one-dimensional chaotic maps are used to generate two chaotic key images needed for circular rotation. A two-dimensional chaotic map is employed for randomly shuffling the coefficients matrix received after mixing. The performance of proposed algorithm is analyzed through experimentation against pixels correlation, peak signal to noise ratio, key sensitivity and key space. It is found that the simulation results validate the effectiveness of the proposed octuple images encryption algorithm.

Keywords Octuple image encryption · Security · Chaotic systems · Circular rotation · Wavelet domain

M. Ahmad (✉) · B. Alam
Department of Computer Engineering, Faculty of Engineering and Technology,
Jamia Millia Islamia, New Delhi 110025, India
e-mail: musheer.cse@gmail.com

A. Jain
Department of Computer and Information Science and Engineering, University of Florida,
Florida 32611, USA

V. Khare
Department of Computer Science and Information Technology, Jaypee Institute of Information
Technology, Noida 201301, India

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_50,

1 Introduction

Images are the most commonly and widely used data in the application areas of defense and military, multimedia broadcasting, tele-medicine, tele-education, weather forecasting, etc. The recent advancements in multimedia/network communication technologies have alleviated and encouraged their unauthorized access and illegal alteration. Consequently, the area of image security has become more challenging and the need of effective, efficient and standardized image encryption methods are growing continuously. To secure and prevent the images sent over the attack-prone networks against intruder's attacks, various optical and digital image encryption techniques have been suggested in the literature [1–13]. Most of these image encryption proposals deal with solving the problem of encrypting a single image. Now, instead of encrypting one image, a number of such images can be encrypted simultaneously which can be transmitted to the receiver. Multiple-image encryption (MIE) has paid attentions of researchers and scholars worldwide, and becomes one of the potential areas of research in the field of information security. To tackle the aforementioned challenge, the researchers and scholars have proposed a number of multiple-image encryption algorithms [9–23]. Most of the existing MIE proposals are based on optical encryption techniques such as wavelength multiplexing [14, 15], random phase matching [16], position multiplexing [17], double random phase encoding [9, 18], multi-channel encryption [19] etc. that utilizes one or other fractional transform(s). Apart from the optical techniques, there are proposals which are based on blind source separation [22] and random grids [23] techniques.

In this paper, a digital image encryption algorithm is proposed which uses a novel approach to encrypt eight different images. Discrete wavelet transform is applied to extract the low frequency components of all images. The low frequency components of images are combined together to get transformed matrices. The chaotic maps are employed in the encryption process, which makes the algorithm more robust and key sensitive. The chaotic key images are used to rotate circularly the values of transformed matrices separately. The rotated transformed matrices are then randomly intermixed to fuse the information of images, followed by its shuffling using random sequences extracted from 2D chaotic map. The remaining of this paper is organized as follows: Section 2 discusses the proposed image encryption algorithm along with the brief illustration of preliminary concepts used in the design. Section 3 discusses the experimental and simulation analyses in detail. Finally, the conclusion of the work is summarized in Sect. 4.

2 Proposed Octuple Image Encryption

The proposed octuple images encryption algorithm has the following subsections.

2.1 PWLCM

The piecewise linear chaotic map (PWLCM) is a dynamical system that exhibits chaotic behavior for all values of parameter $p \in (0, 1)$, defined in Eqn (1) [24]. Where *u(0)* is initial condition, $n \geq 0$ is the number of iterations and *u(n)* ∈ (0, 1) for all *n*. The research shows that the map has largest +ve lyapunov exponent at $p = 0.5$ and the trajectory visit the entire interval [0, 1] for every value of control parameter.

$$u(n+1) = \begin{cases} \dfrac{u(n)}{p} & x(n) \in (0, p] \\ \dfrac{1-u(n)}{1-p} & x(n) \in (p, 1) \end{cases} \tag{1}$$

2.2 Chaotic 1D Logistic Map

The one-dimensional Logistic map proposed by May [25] is one of the simplest nonlinear chaotic discrete systems that exhibit chaotic behavior; it is governed by the following equation.

$$w(n+1) = \lambda w(n)(1 - w(n)) \tag{2}$$

Where $w(0)$ is initial condition, λ is the system parameter and n is the number of iterations. The research shows that the map is chaotic for $3.57 < \lambda < 4$ and *w(n+1)* ∈ (0, 1) for all *n*.

2.3 Chaotic 2D Logistic Map

The two-dimensional coupled Logistic map [26] is described as follows.

$$\begin{aligned} x(i+1) &= \mu_1 x(i)(1 - x(i)) + \gamma_1 y^2(i) \\ y(i+1) &= \mu_2 y(i)(1 - y(i)) + \gamma_2 (x^2(i) + x(i)y(i)) \end{aligned} \tag{3}$$

Three quadratic coupling terms introduce strength to the complexity of 2D Logistic map. This map is chaotic when $2.75 < \mu_1 \leq 3.4$, $2.7 < \mu_2 \leq 3.45$, $0.15 < \gamma_1 \leq 0.21$, $0.13 < \gamma_2 \leq 0.15$ and generate chaotic sequences $x(i)$, $y(i) \in (0, 1)$. The statistical analysis of two sequences generated by the coupled Logistic map shows that they have poor balance, autocorrelation and cross-correlation properties. To improve the statistical properties of the sequences, following pretreatment is done.

$$x(i) = x(i) \times 10^5 - \mathit{floor}(x(i) \times 10^5)\ \&y(i) = y(i) \times 10^5 - \mathit{floor}(y(i) \times 10^5) \tag{4}$$

Now, the preprocessed sequences have better statistical properties and they can be utilized in a cryptographic process. The shuffling operation in the algorithm is performed through sequences generated by the 2D coupled logistic map.

2.4 Proposed Algorithm

The processing steps of the proposed octuple-images encryption algorithm are as follows:

Step 1. Apply discrete wavelet transform at $level = 1$ to all eight images and extracts their low frequency components.
Step 2. Combine four low frequency bands together to obtained two transformed matrices *wcombined1*(i, j) and *wcombined2*(i, j) as shown in Fig. 1.
Step 3. Take initial conditions for chaotic maps (1) to produce chaotic sequence.
Step 4. Process the sequence and generate one chaotic key image shown in Fig. 4a.
Step 5. Circularly rotate the values in each row of *wcombined1*(i, j) by the value equal to *Sum*(i).

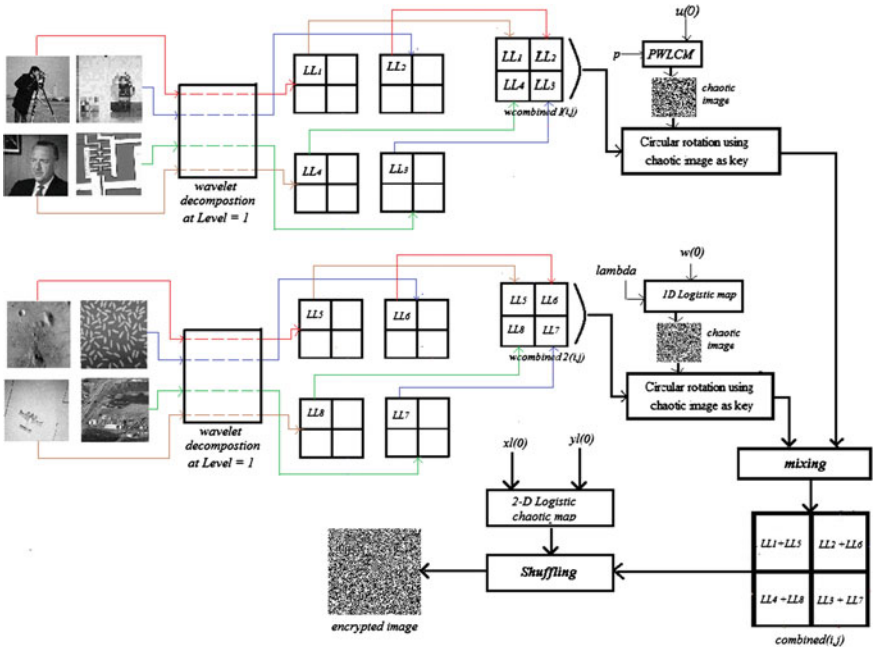


Fig. 1 Block diagram of proposed octuple images encryption algorithm

Step 6. Perform the operations similar to steps 3–5 using chaotic map (2) to generate other chaotic key image shown in Fig. 4b and circularly rotate the values in each column of *wcombined2*(*i, j*) by the value equal to *Sum*(*j*).

Step 7. Combine *wcombined1*(*i, j*) and *wcombined2*(*i, j*) in a fashion shown in Fig. 1. Both *wcombined1*(*i, j*) and *wcombined2*(*i, j*) can be viewed as having four quadrants. The quadrant 1 of *wcombined1*(*i, j*) is combined with quadrant 1 of *wcombined2*(*i, j*), this will mix up the transformed values and results in lowering the correlation between adjacent pixels. The mixing is carried out by inserting the values of *wcombined1*(*i, j*) and *wcombined2*(*i, j*) at every alternate rows and columns of *combined*(*i, j*) matrix.

Step 8. Take proper initial conditions of chaotic map (3). Apply iterations to generate $x(i)$, $y(i)$ sequences and preprocess these sequences through Eqn. 4.

Step 9. Generate the random coordinates for rows and columns as: $X(i)$=1+{*floor* ($x(i)\times 10^{10}$)}mod(2N)&$Y(i)$=1+{*floor*($y(i)\times 10^{10}$)}mod(2N). It is ensure that the elements of $X(i)$ and $Y(i)$ are random indices of rows and columns without any repetition.

Step 10. Now, shuffle the pixel values of *combined*(*i, j*) matrix by relocating them from location (*i, j*) to new random location ($X(i)$, $Y(i)$), where $i, j = 1, 2, \ldots.., 2 \times N$. The resultant is the final encrypted image.

The algorithm is symmetric and deterministic, the decryption process proceeds similar to the process described above but in reverse order. The block diagram of the proposed algorithm is shown below.

3 Simulation Results

The proposed scheme is experimented with eight gray-scale images shown in Fig. 2. The initial values taken for encrypting the images: $u(0) = 0.43456$, $p = 0.22345$, $w(0) = 0.31597$, $\lambda = 3.998$, $x(0) = 0.1354$, $y(0) = 0.5734$, $\mu_1 = 2.93$, $\mu_2 = 3.17$, $\gamma_1 = 0.197$ and $\gamma_2 = 0.139$. The mother wavelet filter used is '*haar*' at level = 1. The two chaotic key images of size 256 × 256 generated by piece-wise linear chaotic map and 1D logistic map are shown in Fig. 4. The encrypted image obtained using proposed algorithm is shown in Fig. 3. As evident from the encrypted image, that it is highly distorted, undistinguishable and appears like a noise image. The decrypted images are shown in Fig. 5. In order to evaluate the encryption and decryption performance of the proposed octuple image encryption algorithm, the simulation analyses like: correlation, key sensitivity, peak signal to noise ratio and key space analysis are performed.

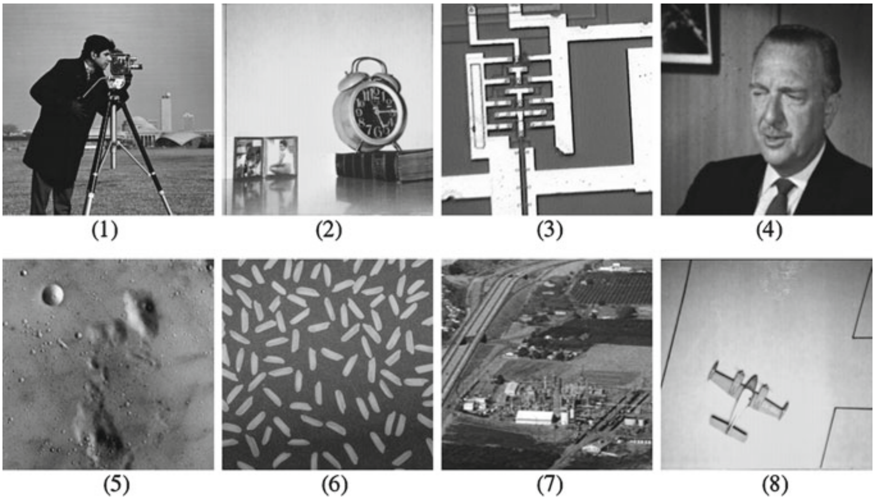


Fig. 2 Eight plain-images each of size 256 × 256

3.1 Pixels Correlation

An image encryption algorithm should be able to completely eliminate the correlation of adjacent pixels in images. The correlation coefficients obtained for encrypted image are listed Table 1. The values show that the proposed algorithm effectively decorrelates the adjacent pixels of encrypted image as the values are very close to 0. Where as, the values listed in Table 2 reveal that the pixels of original images are highly correlated to each other as coefficients are close to 1. The correlation between the original and decrypted images is also evaluated and provided in Table 3. The correlation between the pairs of images as desired is fairly high, indicating that the two images are almost similar to each other.

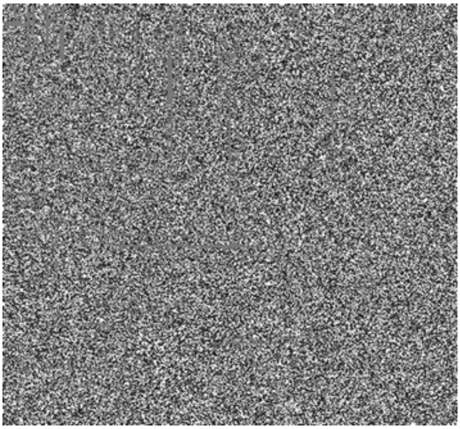


Fig. 3 Resultant encrypted image of size 512 × 512

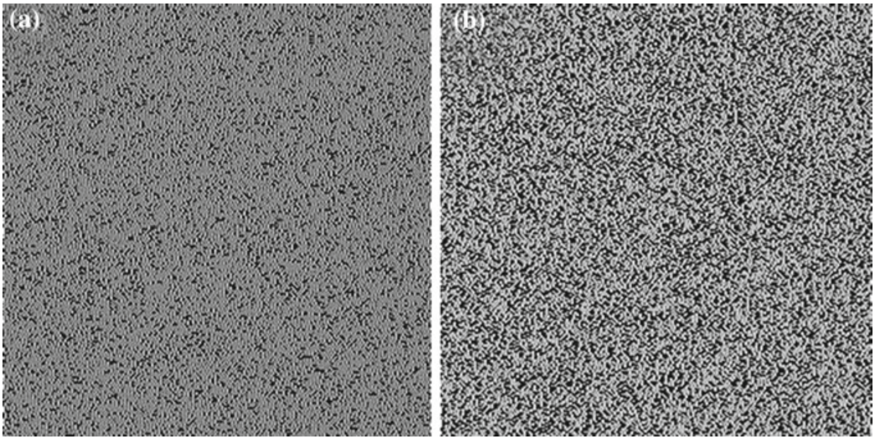


Fig. 4 Chaotic key images using. **a** PWLCM. **b** 1D Logistic map

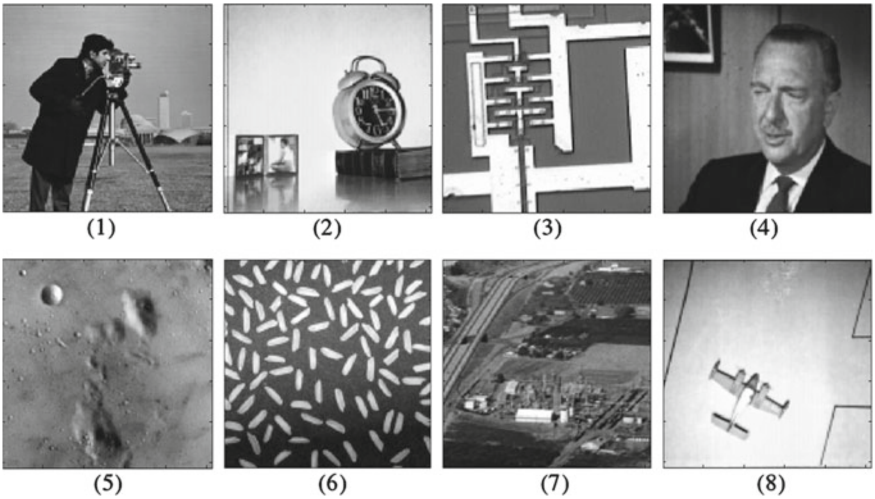


Fig. 5 Eight decrypted images each of size 256 × 256

3.2 Peak Signal to Noise Ratio

Peak signal to noise ratio (PSNR) is an important indicator of amount of distortion taken place while processing the images through encoding/decoding algorithms. The PSNR between the original and decrypted images are evaluated to quantify the loss of information during encryption/decryption. The PSNRs obtained for eight set of

Table 1 Correlation among adjacent pixels of encrypted image

Image	Vertical	Horizontal	Diagonal
Figure 3	−0.070214	−0.078290	0.003924

Table 2 Correlation among adjacent pixels of original images

Image	Vertical	Horizontal	Diagonal
Figure 2 (1)	0.94147	0.96404	0.91506
Figure 2 (2)	0.96001	0.97797	0.94446
Figure 2 (3)	0.98925	0.99169	0.97934
Figure 2 (4)	0.95047	0.90304	0.85820
Figure 2 (5)	0.90945	0.94384	0.91085
Figure 2 (6)	0.90864	0.87070	0.84335
Figure 2 (7)	0.96018	0.91805	0.90977
Figure 2 (8)	0.93659	0.92356	0.88375

Table 3 Correlation and PSNR between original and decrypted images

Images	Corr	PSNR
Figure 2 (1) & 5 (1)	0.976063	34.804
Figure 2 (2) & 5 (2)	0.983841	36.228
Figure 2 (3) & 5 (3)	0.958713	34.457
Figure 2 (4) & 5 (4)	0.994618	38.612
Figure 2 (5) & 5 (5)	0.965018	35.347
Figure 2 (6) & 5 (6)	0.970215	34.189
Figure 2 (7) & 5 (7)	0.967671	33.167
Figure 2 (8) & 5 (8)	0.965439	39.080

images are enumerated in Table 3. PSNR measures indicate that the loss due to lost high frequency components is acceptable as the values are quite high.

3.3 Key Sensitivity

An encryption algorithm should be sensitive to secret key i.e. a small change in secret key during encryption process should results into a completely different encrypted image. In proposed scheme, a tiny change in key even of the order of (Δ=) 10^{-10}, results into completely different encrypted image. The pixels percentage difference between two encrypted images (one shown in Fig. 3 and other obtained by changing only one secret key component keeping other unchanged) is obtained and listed in Table 4. It is clear from the entries that the algorithm is highly sensitive to a small change in secret key.

Table 4 Percentage difference between two encrypted images

Changed key	$u(0) + \Delta$	$p + \Delta$	$w(0) + \Delta$	$x(0) + \Delta$	$y(0) + \Delta$
%age difference	74.16	73.25	71.76	95.88	96.07

3.4 Key Space

Key space is the total number of possible keys that can be taken by an encryption algorithm. The key space of the proposed algorithm comes out as: $(10^{10})^5 \approx 2^{170}$.

4 Conclusion

In this paper, we have proposed a novel encryption algorithm which can encrypt eight different images simultaneously. The low frequency components of images are extracted through wavelet decomposition. These components are combined and processed via circular rotations, intermixing followed by shuffling to produce single encrypted image. The chaotic systems are incorporated to make the algorithm robust, secure and efficient. The contents of all original images can be successfully recovered through decryption process. Loss of information due to high frequency components lost is under acceptable range which is permissible in some applications of image encryption. The experimental and simulation tests have been performed which validate the features and effectiveness of the proposed algorithm.

References

1. Javidi B (2005) Optical and digital techniques for information security. Springer, New York
2. Chang CC, Hwang MS, Chen TS (2001) A new encryption algorithm for image cryptosystems. J Syst Softw 58(2):83–91
3. Zhang L, Liao X, Wang X (2005) An image encryption approach based on chaotic maps. Chaos, Solitons Fractals 24(3):759–765
4. Behnia S, Akhshani A, Mahmodi H, Akhavan A (2008) A novel algorithm for image encryption based on mixture of chaotic maps. Chaos, Solitons Fractals 35(2):408–419
5. Refregier P, Javidi B (1995) Optical image encryption based on input plane and Fourier plane random encoding. Opt Lett 20(7):767–769
6. Singh N, Sinha A (2008) Optical image encryption using fractional Fourier transform and chaos. Opt Lasers Eng 46(2):117–123
7. Hennelly BM, Sheridan JT (2003) Image encryption and the fractional Fourier transform. Optik 114(6):251–265
8. Zhao J, Lu H, Song XJ, Li J, Ma Y (2005) Optical image encryption based on multi-stage fractional Fourier transforms and pixel scrambling technique. Opt Commun 249(4–6):493–499.
9. Situ G, Zhang J (2004) Double random-phase encoding in the Fresnel domain. Opt Lett 29(14):1584–1586

10. Joshi M, Chandrashakher Singh K (2008) Color image encryption and decryption for twin images in fractional Fourier domain. Opt Commun 281(23):5713–5720
11. Singh N, Sinha A (2010) Chaos based multiple image encryption using multiple canonical transforms. Opt Laser Technol 42(5):724–731
12. Liu Z, Dai J, Sun X, Liu S (2009) Triple image encryption scheme in fractional Fourier transform domains. Opt Commun 282(4):518–522
13. Liu Z, Liu S (2007) Double image encryption based on iterative fractional Fourier transform. Opt Commun 275:324–329
14. Situ G, Zhang J (2005) Multiple-image encryption by wavelength multiplexing. Opt Lett 30(11):1306–1308
15. Amaya D, Tebaldi M, Torroba R, Bolognini N (2009) Wavelength multiplexing encryption using joint transform correlator architecture. Appl Opt 48(11):2099–2104
16. He M, Cai L, Liu Q, Wang X, Meng X (2005) Multiple image encryption and watermarking by random phase matching. Opt Commun 247(1–3):29–37
17. Situ G, Zhang J (2006) Position multiplexing for multiple-image encryption. J Opt A 8(5):391–397.
18. Tao R, Xin Y, Wang Y (2007) Double image encryption based on random phase encoding in the fractional Fourier domain. Opt Express 15(24):16067–16079
19. Amaya D, Tebaldi M, Torroba R, Bolognini N (2008) Multi-channeled encryption via a joint transform correlator architecture. Appl Opt 47(31):5903–5907
20. Liang XY, Xin Z, Sheng Y, Yao CY (2010) Multiple-image parallel optical encryption. Opt Commun 283(14):2789–2793
21. Liu Z, Zhang Y, Zhao H, Ahmad MA, Liu S (2011) Optical multi-image encryption based on frequency shift. Optik 122(11):1010–1013
22. Lin QH, Yin FL, Mei TM, Liang H (2008) A blind source separation-based method for multiple images encryption. Image Vis Comput 26(6):788–798
23. Chen T H, Tsao KH, Wei KC (2008) Multiple-image encryption by rotating random grids. In: Eighth international conference on intelligent systems design and applications, 252–256 2008.
24. Li S, Chen G, Mou X (2005) On the dynamical degradation of digital piecewise linear chaotic maps. Int J Bifurcat Chaos 15(10):3119–3151
25. May RM (1967) Simple Mathematical Model with very Complicated Dynamics. Nature 261:459–467
26. Wang XY, Shi QJ (2005) New type crisis, hysteresis and fractal in coupled logistic map. Chin J Appl Mech 23:501–506

Part VII
Workshops: The Fourth International Workshop on Network and Communications Security (NCS 2012)

A Strong PVSS Scheme

Fuyou Miao, Xianchang Du, Wenjing Ruan and Suwan Wang

Abstract A verifiable secret sharing allows the dealer to share a share with some participants and can be verified the correctness. Stadler extended the idea and presented a notion of Public Verifiable Secret Scheme (PVSS), which has the property that any one can verify the validity of the share from the sender, but neglects the deceit of receiver. In this paper, we first give a formal definition of strong PVSS (SPVSS) based on PVSS and propose a specific SPVSS scheme. The scheme solves the cheating problems on both sides, which has strong verification requirements. Finally, we show our scheme satisfies the definition of strong PVSS.

Keywords Secret sharing · Secret security · Publicly verifiable · Diffie-Hellman

1 Introduction

The Secret Sharing (SS) scheme was first introduced by Blakley [1] and Shamir [2] independently in 1979. The SS shows that a secret is divided into n shares shared among n participants by a dealer. In this way, any legal subset of the shareholders (access structure) can recover the secret. SS is a method for the storage of secrets, which can be used to preserve key and avoid attacks from the adversary.

Shamir's [2] scheme is based on the threshold method implemented by interpolating polynomial. The (t, n) threshold scheme has been widely applied in secret sharing. A (t, n) threshold scheme has two properties, that is, (1) any t or more than t shareholders can reconstruct the secret in cooperation; (2) but fewer than t can't

F. Miao (✉) · X. Du · W. Ruan
School of Computer Science, University of Science and Technology,
Hefei, People's Republic of China
e-mail: mfy@ustc.edu.cn

S. Wang
School of Computer Science and Technology, University of Anhui, Hefei, China

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_51,

reconstruct the secret. The (t, n) threshold scheme with polynomial is simple and efficient, whose computational complexity is $o(n\log^2 n)$.

In 1985, Chor [3] extended the notion of secret sharing and proposed a definition of verifiable secret sharing (VSS). VSS was introduced to protect against the cheating of dishonest participants. In contrast with traditional secret sharing, a verifiable protocol was added, which was used to verify the validity of shares. That is because that there are probably some malicious participants or dishonest dealers to send incorrect shares mainly in the process of secret distribution and reconstruction. Feldman [4] and Pederson [5] respectively proposed a VSS scheme based on Shamir's [2] scheme and they can effectively detect cheating of participants or dealers. If the VSS scheme can be verified by any party, it has public verifiability, which is called Public Verifiable Secret Sharing (PVSS) introduced by Stadler [6]. Stadler expressed that not only the participants, but everyone else is also able to verify whether the shares have been correctly distributed. Schoenmakers [7] extended the scheme, he required that in the reconstruction process the participants can provide a proof of correctness for each share. The approach is simple and secure computationally. Subsequently Young-Yung [8] proposed an improvement on [7] and presented a PVSS for discrete logs based on the difficulty in computing discrete logs, compared to [7], which is based on the difficulty in deciding Diffie-Hellman. Later Behnad and Eghlidos [9] introduced a new PVSS. They added two different processes: One is disputation used in the case of a complaint against the dealer. The other one called membership proof is used to authenticate the membership of shareholders. The scheme is simpler and can solve the complaint from the participants and verify its validity.

PVSS has a wide range of applications, such as, electronic voting, software key escrow and revocable electronic cash. In previous PVSS, each participant can verify the correctness of shares from the dealer in the process of distribution. Also in the reconstruction process, a shareholder can verify the validity of a share from another shareholder. But the identity of the share provider is unknowable. The share provided to the third verifier is probably incorrect. For example, assuming that a participant distributes one correct share to the participant, but the participant is a cheater, and he announces that the share from sender is invalid. So it's important to guarantee the correctness of information from the dealer or participant, as to detect the deceiving of both sides. In the electronic voting protocol, a voter is regarded as the dealer and talliers are participants. With the PVSS scheme, a voter must provide public proof, but the tallier is probably a cheater, and he is likely to make a false complaint and provide an invalid share to the third party.

In this paper, we extend the idea of PVSS and give a formal definition of strong PVSS (SPVSS), in which both participants must provide the proof of correctness to the third party. This scheme involves Behnad's [9] scheme as the basis for both sharing and reconstruction. In such a way, we add two processes to the traditional PVSS. One in the distribution process, called Distribution-Check, is used to check the validity of shares from the dealer and cheating by the participant. We will show that by this process, a dealer can not only prove to the third party the correctness of a share sent to the participant who complained, but the participant can also prove to the third party that his complaint is reasonable and the share is not modified. The other one, called Reconstruction-Check, is used to check who is the cheater. We prove that our proposed scheme satisfies the definition of SPVSS and meet higher security requirements.

The security of our scheme is based on the Diffie-Hellman problem which is difficult computationally. In our scheme, once a complaint happens, the cheater can be detected. Compared to Behnad's [9] scheme, this scheme does not make use of membership proof and is simpler and more secure [10, 11].

The rest of the paper is organized as follows: In Sect. 2, we briefly describe the concept and characteristics of PVSS. In Sect. 3, we define new notion of strong PVSS and present our SPVSS scheme. Finally, we discuss the security and performance of the new scheme.

2 Review of PVSS

2.1 The Model of PVSS

We will present an informal description of PVSS in Staler [6]

A PVSS scheme consists of a dealer, n participants, $P_1, \ldots P_n$ and an access structure $A \subseteq 2^{\{1,2\ldots,n\}}$, The access structure is monotonous, which means that if A $\in A$ and A $\subseteq$ B then B $\in A$. S is the secret being shared.

A public encryption function E_i is assigned to each participant P_i, such that only he has the corresponding decryption function D_i. The dealer sends S_i to P_i by calculating $S_i = E_i(s_i), i = 1, 2 \ldots n$ and publishing the encrypted share S_i in the share distribution process. Then the algorithm PubVerify is assigned to verify the validity of the encrypted shares. The algorithm has the property that $\exists u \forall A \in 2^{\{1,2\cdots n\}}$:

$$(PubVerify(\{S_i | i \in A\}) = 1) \Rightarrow \text{Recover}(\{D_i(S_i) | i \in A\}) = u$$

if the dealer is honest, $u = s$. The algorithm can be run by any party. If the secret needs to be recovered, an algorithm Recover will be run, which has the feature that

$$\forall \mathrm{A} \in A : \text{Recover}(\{S_i | i \in \mathrm{A}\}) = s$$

and that for all A$\notin A$ it is computationally infeasible to calculate s from $\{\mathrm{S_i} | \mathrm{i} \in \mathrm{A}\}$.

2.2 The Property of PVSS

Firstly, the PVSS scheme has the general property of a VSS scheme, that is, Verification and Secrecy. The receiver can verify the validity of their received shares. The secrecy means that any group in the access structure should not receive a share and recover the secret.

Secondly, another unique property of PVSS is Publicity. The participant or any other party can verify the validity of other participants (with whom they might be able to recover the secret).

3 Definition of Strong PVSS and a Specific Scheme

A PVSS scheme enables a judge to deal with the conflict. But the identities of participants both are uncertain. In other words, the participants both are not credible. Both sides have the probability of behaving fraudulently. We improve the notion and propose new strong publicly verifiable secret sharing that can ensure a higher level of security.

3.1 The Definition of Strong PVSS

Definition 1 Both participants in a strong PVSS scheme can be verified to demonstrate the validity of data provided by them.

In the SPVSS scheme, we add the verification process, S_i' is share provided by each participant P_i, the SPVSS is described as follow: $PubVerify(\{S_i \&\& S_i' | i \in A\} = 1)$

$$\text{Then}(PubVerify(\{S_i' | i \in A\}) = 1) \Rightarrow \text{Recover}(\{D_i(S_i') | i \in A\}) = u$$

3.2 Strong PVSS Scheme

Notation

Throughout this paper the chosen parameters p and q denote large primes such that q divides $p - 1$. Z_p and Z_q are two different fields. G_p is the unique subgroup of Z_p^* of prime order p, and g denotes a generator of G_p.

SPVSS scheme

This SPVSS scheme is based on Shamir's [2] (t, n) threshold scheme. In our scheme, if the share from the dealer is not approved. In the distribution process, then the Distribution-Check protocol is run by other participants or any other party to check the conflict. Similarly, in the Reconstruction process, if a complaint occurs, the reconstruction-check protocol is used to investigate the complaint and check the cheater.

3.2.1 Distribution

The process consists of two steps:

(1) **Distribution of shares**:

The dealer randomly chooses F polynomial of degree at most $t-1$ with coefficients in Z_q.

$$F(x) = F_0 + F_1 x + \cdots + F_{t-1} x^{t-1}$$

And sets $F_0 = s$ which is the secret. The dealer publishes $C_i = g^{F_i}, i = 0, 1, \cdots, t-1$. The participant j, $j = 1, \cdots, n$ registers g^{a_j} as his public key to the dealer, where $a_j \in Z_q$. Then the dealer also publishes g^d as his public key, where $d \in Z_q$. The dealer publishes the encrypted shares $Y_i = s_i (g^{a_i})^d, i = 1, \cdots, n$, where $s_i = F(i)$, using the public keys of the participants.

(2) **Verification of shares**:

The shareholder i decrypts the Y_i using a_i by computing $s_i = Y_i((g^d)^{a_i})^{-1}$, then, verifies the share s_i by computing $g^{s_i} = \prod_{j=0}^{t-1} C_j^{i^j}$. If the equation does not hold, then the participant complains against the dealer.

3.2.2 Distribution-Check

If the shareholder A complains against the dealer D, the third party R will run to verify the validity of the complaint and vote against the dealer D or the shareholder A by the following protocol. Firstly, we assume that the public secret of A and D is g^a and g^d and the published encrypted share is $Y_A = s_A (g^a)^d$.

(1) **Verification of shareholder**:

1. R chooses randomly $r \in Z_q$, calculates g^r and sends to shareholder A;
2. A calculates $\lambda = (g^r)^{g^{ad}}$ and sends back to R;
3. R then sends λ to the dealer D;
4. D calculates $(g^{ad})^{-1}$ and $\eta = \lambda^{(g^{ad})^{-1}}$, then publish η.
5. Both R and A verify if $\eta = g^r$. If it holds, the protocol is continued. Otherwise, A and D respectively provide the values a and d to R, R verifies the dishonest party by calculating g^a, g^d and λ.
6. R randomly chooses another value $t \in Z_q$, calculates $\sigma = \lambda^{r^{-1} t}$ and sends to A;
7. A calculates σ^{s_A} and sends back to R;
8. R verifies if $(\sigma^{s_A})^{t^{-1}} = g^{Y_A}$, if it holds, it says that the share provided by shareholder is true and it is not modified. Then the protocol checks the dealer's identity.

(2) **Verification of dealer**

9. R chooses $s \in Z_q$, and publishes $\rho = g^{s s_A g^{ab}}$;
10. A and D independently calculate $\delta = \rho^{(g^{ab})^{-1}}$ and send the results to R;
11. R verifies if the two values received are equal, if they are, the protocol is continued. Otherwise, A and D send the values a and d to R to check the cheat;

12. R checks if $\delta^{s^{-1}} = \prod_{i=0}^{t-1} C_i^{j_{A^i}}$, if the equality holds, then s_A is correct, else it is not valid and the dealer is a cheater.

3.3 Reconstruction

The process consists of two steps;

3.3.1 Verification of Share

When at least t shareholders will cooperate to recover the secret, there will be a shareholder sending share to another shareholder,this protocol can be run to verify the correctness of the share.

a and *b* respectively are shareholder A and shareholder B's private key. Shareholder A encrypts the share s_A by calculating $Y_{AB} = s_A g^{ab}$, and publishes the value.

Shareholder B decrypts the encrypted share by calculating $s_A = Y_{AB}(g^{ab})^{-1}$, and verifies if $g^{s_A} = \prod_{i=0}^{t-1} C_i^{j_A^i}$, if it holds, the share provided by shareholder A is correct. Else the following protocol reconstruction-check is run.
Reconstruction-check

1. The shareholder B sends g^{s_A} calculated to R;
2. R sends g^{s_A} to the shareholder A, and A calculates $\alpha = (g^{s_A})^{g^{ab}}$ sends back to R;
3. B then sends $\beta = (g^{s_A})^{g^{ab}}$ to R, R verifies if $\alpha = \beta$, if it holds then the protocol is continued, else A and send a and to R;
4. R verifies if $\alpha = \beta = (g^{s_A})^{g^{ab}}$ if the equation does not hold, then we can be sure B is a cheater, else the protocol is continued;
5. R verifies if $g^{s_A} = \prod_{i=0}^{t-1} C_i^{j_A^i}$, if it does not hold, then A is a cheater;

3.3.2 Recovering the Secret

When all the shares received are correct, then the secret is reconstructed as follows:

$$s = \sum_{i=1}^{t} w_i s_i$$

where $w_i = \prod_{j \neq i} \frac{i}{j-1}$ is a lagrange coefficient.

4 Security

Our scheme adds two new processes to the protocol, Distribution-Check and Reconstruction-Check. The two processes are both used to check which party is the cheater once a complaint happens. This mainly reflects two aspects:

- Participant A is a cheater and sends party B an invalid share, then B reports an error;
- A sends B the correct share , but B is a cheater, then B also reports an error;

Hence, the security of the protocol is based on the security of the two stages. Distribution-check is run when there is a complaint against the dealer.

Lemma 1 The shareholder A cannot commit an invalid and modified share.

Proof By verifying $(\sigma^{s_A})^{t^{-1}} = g^{Y_A}$ in step 8 of the Distribution-Check stage, R is able to believe the share provided by the shareholder is consistent with what was sent by the dealer, and the shareholder does not modify the share. Firstly, if the shareholder provides the invalid $(g^{ab})'$ to replace λ with λ' in step 2. Then during step 5, the value η' received will be $(g^r)^{(g^{ab^{-1}})^{g^{ab'}}}$ and $\eta' \neq \eta$. At this time, A and D have to provide the value a and d to verify the fact. If A modifies the share s_A to s'_A, and it must guarantees $(g^t)^{s_A g^{ab}} = (g^t)^{s'_A g^{ab}}$. Even if A can solve the discrete logarithm, he must compute $ts_A g^{ab}$ from $ts'_A g^{ab}$, which is impossible, because the value t is unknown.

Lemma 2 the dealer cannot send an invalid share to the participant

Proof By verifying $\delta_A = \delta_D$ in step 10, R is convinced that A and D agree on the common key. Due to the random value s and the difficulty of solving the discrete logarithm, the dealer D cannot forge an invalid δ_D to make the equation $\delta_D^{s^{-1}} = \delta^{s^{-1}}$.

Lemma 3 1. The shareholder B cannot send an invalid share from shareholder A; 2. The share received by shareholder B cannot be invalid;

Proof

1. Firstly, the shareholder B cannot send a false value $(g^{ab})'$ such that the equation $(g^{sA})^{g^{ab}} = (g^{sA})^{(g^{ab})'}$ holds. This is a discrete logarithm problem. Similarly, B also cannot forge a value s'_A to deceive R to make $(g^{s_A})^{g^{ab}} = (g^{s'_A})^{g^{ab}}$, this is computationally impossible. Therefore, the shareholder must provide a correct share except that he can solve the discrete logarithm problem.
2. By verifying the equality in step 5, R is convinced that B sent an encrypted correct share. On the other hand, from step 2, R was convinced that A would have the same share value.

5 Conclusion

In this paper, we extend the basic definition of PVSS and give a formal definition of SPVSS. We show that traditional PVSS scheme can't ensure the cheat of both sides. We develope a SPVSS scheme, in which we stress that each party's behave can be tested. If one party is malicious, our scheme can detect it and determine the complaint. Compared to previous schemes, our scheme adds two stages: (1) Distribution-Check, where the dealer and every shareholder can both prove the correctness of the share to other parties; (2) Reconstruction-Check phase, where each shareholder can prove to other parties the correctness of a share to reconstruct the secret. We prove that our scheme satisfy the security of PVSS and have a higher level of security.

Acknowledgments The paper is supported by NSF of China (granted no. 60970128, No. 61170233, No. 61272472), Youth Innovation Foundation of USTC of 2010 and China Postdoctoral Science Foundation (No. 2011M501397).

References

1. Blakley GR (1979) Safeguarding cryptographic keys. In: Proceedings of the national computer conference 1979, vol 48. American Federation of Information Processing Societies, pp 313–317
2. Shamir A (1979) How to share a secret. Commun ACM 22(11):612–613
3. Chor B, Goldwasser S, Micali S, Awerbuch B (1985) Verifiable secret sharing and achieving simultaneity in the presence of faults. In: 26th annual symposium on foundation of computer, science, pp 383–395
4. Feldman P (1987) A practical scheme for non-interactive verifiable secret sharing. In: Proceedings of the 28th IEEE symposium on foundations of computer science, pp 427–437
5. Pederson TP (1992) Non-interactive and information-theotetic secure verifiable secret sharing. In: Advances in cryptology-CRYPTO'91, LNVS576, pp 129–140
6. Stadler M (1996) Publicly verifiable secret sharing. In: Advance in cryptology-EUROCRYPT'96. LNCS 1070:190–199
7. Schoenmakers B (1999) A simple publicly verifiable secret sharing scheme and its application to electronic voting. In: CRYPTO'99. LNCS 1666:148–164
8. Young A, Yung M (2001) A PVSS as hard as discrete log and shareholder separability. In: PKC 2001, LNCS 1992, 2001, pp 287–299
9. Behnad A, Eghlidos T (2008) A new publicly verifiable secret sharing scheme. Scientia Iranica 15(2):246–251
10. Fujisaki E, Okamoto T (1998) A practical and provably secure scheme for publicly verifiable secret sharing and its applications. In: Eurocrypt'98, LNCS 1403, 1998, pp 32–46
11. Chor B, Goldwasser S, Micali S, Awerbuch B (1985) Verifiable secret sharing and achieving simultaneity in the presence of faults. In: Proceedings of 26th annual symposium on foundations of computer science 1985, pp 383–395

Secure Cosine Similarity Computation with Malicious Adversaries

Dexin Yang, Baolin Xu, Bo Yang and Jianping Wang

Abstract Similarity coefficients play an important role in many aspects. Recently, several schemes were proposed, but these schemes aimed to compute the similarity coefficients of binary data. In this paper, a novel scheme which can compute the coefficients of integer is proposed. To the best knowledge of us, this is the first scheme which can resist malicious adversaries attack.

Keywords Similarity coefficients · Distributed EIGamal encryption · Zero-knowledge proof · Secure two-party computation

1 Introduction

Cosine similarity is a measure of similarity between two vectors by measuring the cosine of the angle between them. The cosine of 0 is 1, and less than 1 for any other angle; the lowest value of the cosine is -1. The cosine of the angle between two vectors thus determines whether two vectors are pointing in roughly the same direction (https://en.wikipedia.org/wiki/Cosine-similarity). Many application domains need

D. Yang · J. Wang
Department of Information, Guangzhou City Polytechnic, Guangzhou 510405, China
e-mail: yangdexin@21cn.com

J. Wang
e-mail: wjp@gcp.edu.cn

B. Xu
Department of Information, Guangdong Baiyun Institute, Guangzhou 510460, China
e-mail: 573260286@qq.com

B. Yang (✉)
School of Computer Science, Shaanxi Normal University, 710062 Xi'an, China
e-mail: byang@snnu.edu.cn

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_52,

this parameter to analyze data, such as privacy-preserving data mining, biometric matching etc.

The functionality of the privacy-preserving cosine similarity for integer data (Denoted by $\mathscr{F}_{\mathscr{C}\mathscr{C}}$) can be described as follows. Consider P_1 has a vector $\mathbf{a} = (\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_\mathrm{n})$, P_2 has a vector $\mathbf{b} = (\mathrm{b}_1, \mathrm{b}_2, \ldots, \mathrm{b}_\mathrm{n})$, where $a_i, b_i \in Z_p$. After the computation P_1 gets the result the cosine correlative coefficient SC and P_2 gets nothing.

Related works. Secure two-party computation allows two parties to jointly compute some functions with their private inputs, while preserving the privacy of two parties private inputs. Research on the general functionality of secure computation was first proposed in [1] in the semi-honest model. Lately, Goldreich [2], Malkhi [3], Lindell and Pinkas [4, 5] extended in the presence of malicious adversaries. Even though the general solution of secure multiparty computations has given by Goldreich [6]. However, these general solutions are inefficient for practical uses, because these protocols were constructed based on the boolean circuit or the arithmetic circuit of the functionality. When the circuit of the functionality became complex enough, the complexity of this protocol will be too lower to tolerate. Till now, the protocol which can resist to the attacks of malicious adversaries were the focus works of cryptographers. Therefore, it is necessary to construct the protocol which can compute cosine correlative coefficient of two vectors in the malicious model.

Kikuchi et al. [7] gave the first protocol to compute two vectors cosine correlative coefficient based on zero knowledge proof of range and applied this protocol to biometric authentication. This protocol is based on zero-knowledge proofs and Fujisaki-Okamoto commitments [8]. Recently, Wong et al. [9] proposed a new protocol which can compute the similarity coefficient of two binary vectors in the presence of malicious adversaries. Later, Zhang et al. [10] pointed out this scheme is not secure, and another scheme which can overcome the shortage of Wong's scheme is proposed.

Our results. In this paper, a new protocol which can compute the cosine correlative coefficient is proposed. Our protocol can resist the attacks of malicious adversaries, and we give the standard simulation-based security proof.

Our main technical tools include distributed EIGamal encryption [11] and zero-knowledge proofs of knowledge. The main property of distributed EIGamal encryption is that the parties must cooperate while in decrypting stage because each party has partial decrypt key.

2 Preliminaries

2.1 Cosine Correlative Similarity

Let $\mathbf{a} = (\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_\mathrm{n})$, $\mathbf{b} = (\mathrm{b}_1, \mathrm{b}_2, \ldots, \mathrm{b}_\mathrm{n})$ be two n-dimensional integer vectors. We consider the cosine similarity between **a** and **b**,which will be evaluated in privacy-preserving in later section.

Definition 1 *A cosine correlation is a similarity between* **a** *and* **b** *defined as*

$$cos(\mathbf{a},\mathbf{b}) = \frac{\mathbf{a}\cdot\mathbf{b}}{\|\mathbf{a}\|\cdot\|\mathbf{b}\|} = \frac{\mathrm{a_1b_1+\cdots+a_nb_n}}{\sqrt{\mathrm{a_1^2+\cdots+a_n^2}}\sqrt{\mathrm{b_1^2+\cdots+b_n^2}}}$$

For normalization $\mathbf{a}, \mathbf{b}(\|\mathbf{a}\| = 1, \|\mathbf{b}\| = 1)$, the cosine correlation can be simplified as

$$cos(\mathbf{a},\mathbf{b}) = \mathbf{a}\cdot\mathbf{b} = a_1b_1 + \cdots + a_nb_n$$

where $\|\mathbf{a}\|$ is a norm of **a**.

The proposed scheme in this paper is focus on computing the cosine correlation coefficient of two normalization vectors **a** and **b**.

2.2 Distributed ElGamal Encryption

ElGamal encryption [12] is a probabilistic and homomorphic public-key crypto system. Let p and q be two large primes such that q divides $p-1$. G_q denotes Z_p^* unique multiplicative subgroup of order q. All computations in the remainder of this paper are modulo p unless otherwise noted. The private key is $x \in Z_q$, and the public key is $y = g^x$ ($g \in G_q$ is a generator). A message $m \in G_q$ is encrypted by computing the ciphertext tuple

$$(\alpha, \beta) = (my^r, g^r)$$

where r is an arbitrary random number in Z_q, chosen by the encrypter.

A message is decrypted by computing

$$\frac{\alpha}{\beta^x} = \frac{my^r}{(g^r)^x} = m$$

ElGamal is homomorphic as the component-wise product of two ciphertexts

$$(\alpha\alpha', \beta\beta') = (mm'y^{r+r'}, g^{r+r'})$$

represents an encryption of the plaintexts product mm'.

A distributed EIGamal Encryption system [11] is a public-key cryptosysytem which key generation algorithm [13] and decryption algorithm is as follows:

Distributes key generation: Each participant chooses x_i at random and publishes $y_i = g^{x_i}$ along with a zero-knowledge proof of knowledge of y_i's discrete logarithm. The public key is $y = \Pi_{i=1}^n y_i$, the private key is $x = \Sigma_{i=1}^n x_i$. This requires n multiplications, but the computational cost of multiplications is usually negligible in contrast to exponentiations.

Distributed decryption: Given an encrypted message (α, β), each participant publishes $\beta_i = \beta^{x_i}$ and proves its correctness by showing the equality of logarithms of y_i and β_i. The plaintext can be derived by computing $\frac{\alpha}{\Pi_{i=1}^{n}\beta_i}$. Like key generation, decryption can be performed in a constant number of rounds, requiring n multiplications and one exponentiation.

Same as Bo Zhang et al. [10], we also use an additively homomorphic variation of EIGamal Encryption with distributed decryption over a group $\mathbb{G}_q$ in which DDH is hard, i.e., $E_{pk}(m, r) = (g^r, g^m h^r)$.

2.3 Zero Knowledge Proof

In order to obtain security against malicious adversaries, the participants are required to prove the correctness of each protocol step. Zero knowledge proof is a primitive in cryptography. In fact, the proposed protocols can be proven correct by only using Σ-protocols. A Σ-protocol is a three move interaction protocol. In this paper, there are four Σ-protocols are used as follows. We denote these associated functionalities by $\mathscr{F}_{DL}, \mathscr{F}_{EqDL}, \mathscr{F}_{KeyGen}, \mathscr{F}_{IsCipher}$. Next, we simply describe the associated zero-knowledge protocols: $\pi_{DL}, \pi_{EqDL}, \pi_{KeyGen}, \pi_{IsCipher}$.

π_{DL}. The prover can prove to the verifier that he knows the knowledge of the solution x to a discrete logarithm.

$$R_{DL} = \{((G_q, q, g, h), x) | h = g^x\}$$

π_{EqDL}. The prover can prove to the verifier that the solutions of two discrete logarithm problems are equal.

$$R_{EqDL} = \{((G_q, q, g, g_1, g_2, g_3), x) | g_1 = g^x \wedge g_3 = g_2^x\}.$$

π_{KeyGen}. The prover can prove to the verifier that the generation of EIGamal encryption is valid.

$$R_{KeyGen} = \{((G_q, q, g), s_1, s_2) | h = g^{s_1 + s_2}\}$$

$\pi_{IsCipher}$. The prover can prove to the verifier that the ciphertext of EIGamal encryption is valid

$$R_{IsCipher} = \{(G_q, q, g, h), m) | (c_1 = g^r \wedge c_2 = g^m h^r)\}$$

3 The Proposed Scheme

In this section, we give out the protocol (Π_{SC}) which computes the coefficient of two integer vectors in the presence of malicious adversaries. The ideal functionality of coefficient $\mathscr{F}_{SC}$ is as follows:

$$((a_1, a_2, \dots, a_n), (b_1, b_2, \dots, b_n)) \mapsto (SC, \lambda)$$

where λ denotes P_2 gets nothing after the protocol execution, SC denotes the cosine coefficient between two vectors **a**, **b**. In the ideal model, P_1 sends his private input **a** to the third trusted party (TTP), similarly P_2 sends his private input **a** to TTP. Finally, TTP sends SC back to P_1, and nothing to P_2.

The building blocks of our protocol include distributed EIGamal encryption and zero-knowledge proofs. The reason we choose distributed EIGamal encryption rather than original EIGamal encryption is the distributed EIGamal encryption is less complexity in protocol.

The protocol (Π_{SC}) is as follows:

-**Inputs**: The input of P_1 is a n dimensional integer vector $\mathbf{a} = (\mathrm{a}_1, \mathrm{a}_2, \dots, \mathrm{a}_\mathrm{n})$. Similarly, P_2's input is $\mathbf{b} = (\mathrm{b}_1, \mathrm{b}_2, \dots, \mathrm{b}_\mathrm{n})$.

-**Auxiliary Inputs**: Both parties have the security parameter 1^k.

-**The protocol**:

1. P_1, P_2 engage in the protocol $\pi_{KeyGen}(1^k, 1^k)$ to generate the public key $pk = (G_q, g, h = g^{s_1+s_2})$, and the private key (s_1, s_2), shared by P_1, P_2 respectively.
2. P_2 computes: $c' = \|\mathbf{b}\|, \mathrm{E_i} = (\mathrm{g}^{\alpha_\mathrm{i}}, \mathrm{g}^{\mathrm{b_i}/\mathrm{c}'}\mathrm{h}^{\alpha_\mathrm{i}}), \mathrm{i} \in \{1, 2, \dots, \mathrm{n}\}$, and sends E_i to P_1. The parties run the zero-knowledge proof of knowledge $\pi_{iscipher}$, allowing P_1 to verify that the ciphertext E_i is valid.
3. Upon receiving the E_i from P_2, the P_1 computes: $c = \|\mathbf{a}\|, \forall a_i i \in \{1, 2, \dots, n\}$, $G_i = E_i^{a_i/c} = (g^{\alpha_i a_i/c}, g^{a_i b_i/c} h^{\alpha_i a_i/c})$, $E_{SC} = \Pi_{i=1}^n G_i = (g^{\Sigma_{i=1}^n(\alpha_i a_i/c)}, g^{\Sigma_{i=1}^n(a_i b_i/cc')} h^{\Sigma_{i=1}^n(\alpha_i a_i/c)}) = (g^\beta, g^{SC} h^\beta)$ where $\beta = \Sigma_{i=1}^n(\alpha_i a_i/c), SC = \Sigma_{i=1}^n(a_i b_i/cc')$. and sends E_{SC} to party P_2. The parties run the zero-knowledge proof of knowledge $\pi_{iscipher}$, allowing P_2 to verify that the ciphertext E_{SC} is valid.
4. Upon receiving the (g^β, π_{DL}) from P_1, P_2 computes C_1 using his private key s_2 as: $C_1 = (g^\beta)^{s_2}$, and send C_1 to P_1. The parties run the zero-knowledge proof of knowledge , π_{EqDL}, allowing P_1 to verify that C_1 is valid.
5. Upon receiving the C_1 from P_1, P_1 decrypts and obtains g^{SC}, where SC is the cosine coefficient of the two vectors **a**, **b**. At last, P_1 evaluates SC as follows.

   (a) If $g^{SC} = 1$, then $SC = 0$;
   (b) If $g^{SC} = g$, then $SC = 1$.

4 Security Analysis

Theorem 1 *Assume that $\pi_{DL}, \pi_{EqDL}, \pi_{KeyGen}, \pi_{IsCipher}$ are as described in Sect. 2 and that (Gen, E, D) is the ElGamal scheme. The Π_{SC} correctly evaluates the cosine coefficient of two n-dimension variables in the presence of malicious adversaries.*

The proof of this theorem 1 is obviously.

Theorem 2 *Assume that $\pi_{DL}, \pi_{EqDL}, \pi_{KeyGen}, \pi_{IsCipher}$ are as described in Sect. 2 and that (Gen, E, D) is the ElGamal scheme. The Π_{SC} securely evaluates the cosine coefficient of two n-dimension variables in the presence of malicious adversaries.*

Proof We prove this theorem in the hybrid model, where a third trusted party is introduced to compute the ideal functionality $\mathscr{F}_{DL}, \mathscr{F}_{EqDL}, \mathscr{F}_{KeyGen}, \mathscr{F}_{IsCipher}$. As usual, we analyze two cases as P_1 is corrupted and P_2 is corrupted separately.

P_1 **is corrupted**. Assume that P_1 is corrupted by adversary $\mathscr{A}$ with the auxiliary input z in the real model. We construct a simulator $\mathscr{S}$, who runs in the ideal model with the third trusted party computing the functionality F_{D_C}. S works as follows.

1. $\mathscr{S}$ is given $\mathscr{A}$'s input and auxiliary input, and invokes $\mathscr{A}$ on these values.
2. $\mathscr{S}$ first emulates the trusted party for π_{KeyGen} as follows. It first two random elements $s_1, s_2 \in \mathbb{Z}_q$, and hands $\mathscr{A} s_1$ and the public key $(\mathbb{G}_{||}, q, g, h = g^{s_1+s_2})$.
3. $\mathscr{S}$ receives from p_2 n encryptions and P_2's input for the trusted party for $F_{iscipher}$, then define $\mathscr{A}$'s inputs as $\mathbf{b}$.
4. Then $\mathscr{S}$ sends $\mathbf{b}$ to the trusted party to compute F_{SC} to complete the simulation in the ideal model. Let I_{D_C} be the returned value from the trusted party.
5. Next $\mathscr{S}$ randomly chooses $\mathbf{a}' = (a'_1, a'_2, \ldots, a'_n)$ conditioned on that the cosine coefficient equals to I_{D_C}. $\mathscr{S}$ completes the execution as the honest party P_2 would on inputs $\mathbf{a}'$.
6. If at any step, $\mathscr{A}$ sends an invalid message, $\mathscr{S}$ aborts sends $\perp$ to the trusted party for F_{D_C}. Otherwise it outputs whatever $\mathscr{S}$ does.

The difference between the above simulation and the real hybrid model is that $\mathscr{S}$ who does not have the real P_1's input $\mathbf{a}$, simulates following steps with the randomly chosen $\mathbf{a}'$ under the condition that the output of them are the same. The computationally distinguishability of them can be deduced from the sematic security of ElGamal encryption. In other words, if $\mathscr{A}$ can distinguish the simulation from the real execution, we can construct a distinguisher $\mathscr{D}$ to attack the semantic security of ElGamal encryption.

$\mathbf{P_2}$ **is corrupted**. The proof of this part is similar with above. We construct a simulator $\mathscr{S}$ in the ideal model, based on the real adversary $\mathscr{A}$ in the real model. S works as follows.

1. $\mathcal{S}$ is given $\mathcal{A}$'s input and auxiliary input, and invokes $\mathcal{A}$ on these values.
2. $\mathcal{S}$ first emulates the trusted party for π_{KeyGen} as follows. It first two random elements $s_1, s_2 \in \mathbb{Z}_q$, and hands $\mathcal{A} s_1$ and the public key $(\mathbb{G}_{||}, q, g, h = g^{s_1+s_2})$.
3. $\mathcal{S}$ randomly chooses $\mathbf{b}' = (b'_1, 2'_2, \ldots, b'_n)$, then encrypts them using the public key.
4. Next, $\mathcal{S}$ sends the ciphertexts to $\mathcal{A}$, and proves to $\mathcal{A}$ that all the ciphertexts is valid using $\pi_{iscipher}$.
5. $\mathcal{S}$ receives from $\mathcal{A}$ n ciphertexts and $\mathcal{A}$'s input to the trusted party for $F_{iscipher}$, then defines $\mathcal{A}$'s inputs as $\mathbf{b}'$.
6. The $\mathcal{S}$ completes the next step as the honest P_1.
7. If at any step, $\mathcal{A}$ sends an invalid message, $\mathcal{S}$ aborts sends $\perp$ to the trusted party for F_{D_C}. Otherwise $\mathcal{S}$ sends $\mathbf{b}'$ to the trusted party computing F_{SC}, and outputs whatever $\mathcal{S}$ does.

Similar to the case P_1 is corrupted, the difference between the simulation and the real model is that $\mathcal{S}$ uses $\mathbf{b}'$ as P_2's input. However, $\mathbf{b}'$ is encrypted by the public key of a semantic security EIGamal encryption. Same as the above, the analysis of this simulation distribution can be assured by the definition of zero-knowledge proof and semantic security of a public-key encryption.

In summary, we complete the proof of Π_{SC} in the presence of malicious adversaries.

5 Conclusion

Similarity coefficients (also known as coefficients of association) are important measurement techniques used to quantify the extent to which objects resemble one another. There are various similarity coefficients which can be used in different fields. Cosine similarity is a measure of similarity between two vectors by measuring the cosine of the angle between them. In this paper, a new scheme which can compute the cosine correlative of two integer vectors in the presence of malicious adversaries.

Acknowledgments This work is supported by the National Natural Science Foundation of China under Grant 60973134, 61173164, and the Natural Science Foundation of Guangdong Province under Grants 10351806001000000.

References

1. Yao AC-C (1986) How to generate and exchange secrets. In: Proceedings of the 27th annual symposium on foundations of computer science, SFCS '86. IEEE Computer Society, Washington, pp 162–167
2. Goldreich O, Micali S, Wigderson A (1987) How to play any mental game. In: Proceedings of the nineteenth annual ACM symposium on theory of computing, STOC '87. ACM, New York, pp 218–229
3. Malkhi D, Nisan N, Pinkas B, Sella Y (2004) Fairplay—a secure two-party computation system. In: Proceedings of the 13th conference on USENIX security symposium, vol 13, SSYM'04. USENIX Association, Berkeley, pp 20–20
4. Lindell Y, Pinkas B (2007) An efficient protocol for secure two-party computation in the presence of malicious adversaries. In: Proceedings of the 26th annual international conference on advances in cryptology, EUROCRYPT '07, Berlin. Springer, Heidelberg, pp 52–78
5. Lindell Y, Pinkas B (2011) Secure two-party computation via cut-and-choose oblivious transfer. In: Proceedings of the 8th conference on theory of cryptography, TCC'11, Berlin. Springer, Heidelberg, pp 329–346
6. Goldreich O (1998) Secure multi-party computation (working draft). http://citeseer.ist.psu.edu/goldreich98secure.html
7. Kikuchi H, Nagai K, Ogata W, Nishigaki M (2009) Privacy-preserving similarity evaluation and application to remote biometrics authentication. Soft Comput 14(5):529–536
8. Fujisaki E, Okamoto T (1997) Statistical zero knowledge protocols to prove modular polynomial relations. In: Proceedings of the 17th annual international cryptology conference on advances in cryptology, CRYPTO '97. Springer, London, pp 16–30
9. Wong K-S, Kim MH (2012) Privacy-preserving similarity coefficients for binary data. Comput Math Appl. doi:10.1016/j.camwa.2012.02.028
10. Zhang B, Zhang F (2012) Secure similarity coefficients computation with malicious adversaries. Cryptology ePrint Archive, Report 2012/202. http://eprint.iacr.org/
11. Brandt F (2006) Efficient cryptographic protocol design based on distributed el gamal encryption. In: Proceedings of the 8th international conference on information security and cryptology, ICISC'05, Berlin, 2006. Springer, Heidelberg, pp 32–47
12. El Gamal T (1985) A public key cryptosystem and a signature scheme based on discrete logarithms. In: Proceedings of CRYPTO 84 on advances in cryptology. Springer, New York, pp 10–18
13. Pedersen TP (1992) Non-interactive and information-theoretic secure verifiable secret sharing. In: Proceedings of the 11th annual international cryptology conference on advances in cryptology, CRYPTO '91. Springer, London, pp 129–140

Test Suite for Intrusion Detection by Layered Conditional Random Fields Using Mobile Phones

M. Arpitha, V. Geetha, K. H. Gowranga and R. Bhakthavathsalam

Abstract There is high demand to reduce the threat level in networks to ensure the data and services offered by them to be more secure. With the ever increasing number and diverse type of attacks, including new and previously unseen attacks, the effectiveness of an Intrusion Detection System is very important. Earlier works deal with the layered approach and conditional random fields (CRFs) for improving the efficiency and accuracy of an intrusion detection system. In this paper we developed an effective test suite using the layered CRFs. We set up different types of checks at multiple levels in each layer. Our framework examines various attributes at every layer in order to effectively identify any breach of security. Once the attack is detected, it is intimated through mobile phone to the system administrator for safe guarding the server system. We establish experimentally that the layered CRFs can be very effective in detecting intrusions when compared with the previously known techniques.

Keywords Network security · Intrusion detection · Layered approach · Conditional random fields · Mobile phones

M. Arpitha · V. Geetha
Department of Information Science and Engineering, Alpha College of Engineering,
Bangalore, India
e-mail: arpitha119@gmail.com

V. Geetha
e-mail: geethaanjali78@gmail.com

K. H. Gowranga · R. Bhakthavathsalam (✉)
Supercomputer Education and Research Center, Indian Institute of Science, Bangalore, India
e-mail: bhaktha@serc.iisc.ernet.in

K. H. Gowranga
e-mail: gowranga@serc.iisc.ernet.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_53,

1 Introduction

By and large networks are vulnerable to increasing number of attacks by intrusion. Thus securing a network from unwanted malicious traffic is of prime concern. A computer network needs to provide continuous services, such as e-mail, to users, while on the other it stores huge amount of data which is of vital significance. Recently, there has been increasing concern over safeguarding the vast amount of data stored in a network from malicious modifications and disclosure to unauthorized individuals. Intrusion Detection Systems (IDS) are based on two concepts: (a) matching of the previously found and hence known anomalous patterns from an internal database of signatures or (b) building profiles based on normal data and detecting deviations from the expected behaviour [1]. Based on the mode of deployment the Intrusion Detection Systems are classified as Network based and Host based [2]. Network based systems make a decision by analyzing the network logs and packet headers from the incoming and outgoing packets. Host based systems monitor individual systems and use system logs extensively to make any decision. Intrusion Detection Systems are either Signature based or Behaviour based. The Signature based systems build a model based on the available knowledge of the attacks and extract out signatures which are used to build a classifier to detect same or similar patterns. This is also known as Misuse Detection. Complementary to these are the Behaviour based systems which build a model based on the available knowledge of the normal use of the system. The Signature based systems though have very high detection accuracy but they fail when an attack is previously unseen. On the other hand, Behaviour based IDS may have the ability to detect new unseen attacks but have the problem of low detection accuracy. In this paper we proposed a simple yet practical layered approach to intrusion detection/prevention and discussed its various advantages with regards to accuracy and computation. Such a system would be less computational intensive and more accurate. We are currently evaluating different layers individually. We plan to implement this framework as a single system. We propose and evaluate the use of the CRFs [3] also which is a novel technique for the task of Intrusion Detection, and experimentally show that they have higher detection accuracy when compared to any other technique for the same task. We establish that the CRFs perform better than the other methods and offer features which are inherent to the task of Intrusion Detection. Further, our system can be used as a standalone system monitoring an entire Network or a single Host or even a single Application running on a particular host.

1.1 Intrusion Detection

Intrusion detection is the process of monitoring computers or networks for unauthorized entrance, activity, or file modification. An IDS (Intrusion Detection System) is a device or application used to inspect all network traffic, thereby detecting if a system is being targeted by a network attack such as a denial of service attack. In some cases the IDS may also respond to anomalous or malicious traffic by taking action such as blocking the user or source IP address from accessing the network. IDS protects a network and attempt to prevent intrusions, while IDS tools detect whether the network is under attack or not. IDS tools thus form an integral part of a thorough and complete security system. They don't fully guarantee security, but when used with security policy, vulnerability assessments, data encryption, user authentication, access control, and firewalls, they can greatly enhance network safety [4]. Intrusion detection systems serve three essential security functions: they monitor, detect, and respond to unauthorized activity by company insiders and outsider intrusion.

Intrusion detection systems use policies to define certain events that, if detected will issue an alert. Certain intrusion detection systems have the capability of sending out alerts, so that the administrator of the IDS will receive a notification of a possible security incident. Many intrusion detection systems not only recognize a particular incident and issue an appropriate alert, they also respond automatically to the event. Such a response might include logging off a user, disabling a user account, and launching of scripts.

2 New Scheme for Robust IDS

Intrusion detection as a discipline is fairly immature. Commercially available examples of successful intrusion detection systems are limited, although the state of the art is progressing rapidly. However, as new approaches to intrusion detection are introduced, there is one question that seems to emerge continuously: can intrusion detection system detect all illegal requests from unknown computers and will the system maintain accuracy while finding out intrusion, even if it's accurate will it perform with higher efficiency which is always preferred. But none of the existing intrusion detection system will provide both the features of accuracy and efficiency both at once. The main problem with IDS is that the system is not robust. To overcome this drawback we use layered approach with conditional random fields in our paper. The whole concept of our paper is to build an intrusion detection system which is very accurate in detection of request from unknown computers and which is very fast to respond to such intrusions taking place in system which gives efficiency to the system an intimating the administrator about the intrusions through the mobile phone. To do so we have integrated the properties of conditional random fields and the layered approach.

2.1 Existing System

There are a number of methods and frameworks have been proposed and many systems have been built to detect intrusions. Various techniques such as association rules [5], clustering, naive Bayes classifier, support vector machines, genetic algorithms, artificial neural networks, and others have been applied to detect intrusions. These existing systems suffer from a wide range of problems.

a. The features are limited to the entry level of the packets and require the no. of records to be large. They tend to produce a large number of rules that increases the system's complexity.
b. Some methods consider the features independently and are unable to capture the relationship between different features of a single record. This further degrades the attack detection strength of the system.
c. Some existing systems are attack specific and hence they would build networks which rapidly increases as the detection load increases.

All these tend to decrease the efficiency and accuracy of the system.

2.2 Proposed System

In our proposed system we describe the Layer-based Intrusion Detection System (LIDS) [6, 7]. The LIDS draws its motivation from what we call as the Airport Security model, where a number of security checks are performed one after the other in a sequence. Similar to this model, the LIDS represents a sequential Layered Approach and is based on ensuring availability, confidentiality, and integrity of data and (or) services over a network [8].

The goal of using a layered model is to reduce computation and the overall time required to detect anomalous events. The time required to detect an intrusive event is significant and can be reduced by eliminating the communication overhead among different layers. Every layer in the LIDS framework is trained separately and then deployed sequentially. We define four layers they are Probe layer, DoS layer, R2L layer, and U2R layer. Each layer is separately trained with a small set of features. Feature selection is significant for Layered Approach. The layers essentially act as filters that block any anomalous connection, thereby eliminating the need of further processing at subsequent layers enabling quick response to intrusion.

The effect of such a sequence of layers is that the anomalous events are identified and blocked as soon as they are detected. Once the attack is detected, it is intimated through mobile phone to the system administrator for safe guarding the server system. We implement the LIDS and select a small set of features for every layer rather than using all the 41 features. This results in significant performance improvement during both the training and the testing of the system.

In many situations, there is a trade-off between efficiency and accuracy of the system and there can be various avenues to improve system performance. Methods such as naive Bayes assume independence among the observed data [9]. This certainly increases system efficiency, but it may severely affect the accuracy. To balance this trade-off, we use the CRFs that are more accurate, though expensive, but we implement the Layered Approach to improve overall system performance (Fig. 1).

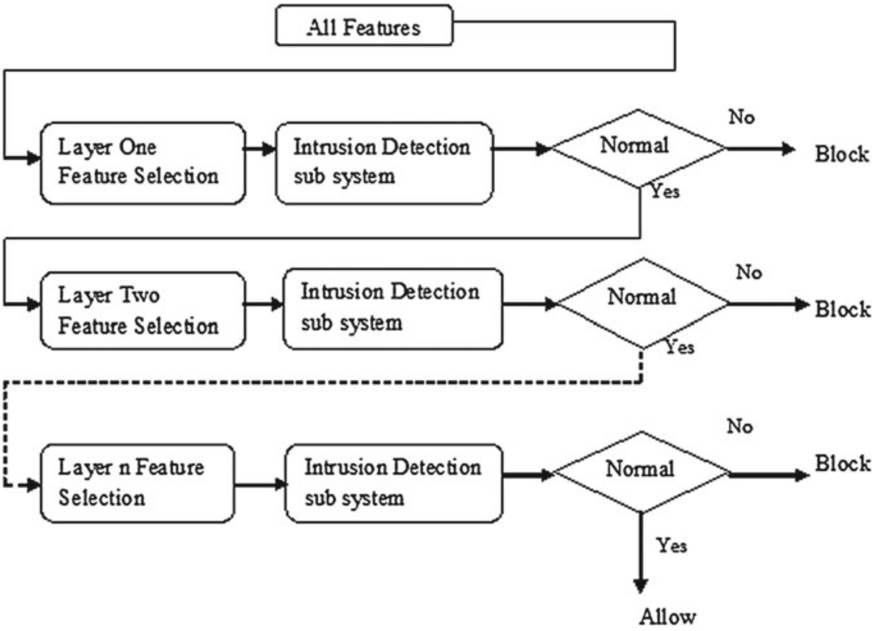

All Features
Layer One Feature Selection
Intrusion Detection sub system
Normal
No
Block
Yes
Layer Two Feature Selection
Intrusion Detection sub system
Normal
No
Block
Yes
Layer n Feature Selection
Intrusion Detection sub system
Normal
No
Block
Yes
Allow

Fig. 1 Proposed system

The performance of our proposed system, Layered CRFs, is comparable to that of the decision trees and the Naive Bayes, and our system has higher attack detection accuracy. Our proposed system, Layered CRFs, performs significantly better than other systems. Other approaches for detecting intrusion include the use of autonomous and probabilistic agents for intrusion detection. These methods are generally aimed at developing a distributed intrusion detection system.

3 Implementation

Implementation is the stage when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage comprises of careful planning and investigation of the existing system. It also deals with the constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

3.1 Layered Approach for Intrusion Detection

Layer-based Intrusion Detection System (LIDS) derives its motivation from the Airport Security model, where a number of security checks are performed one after

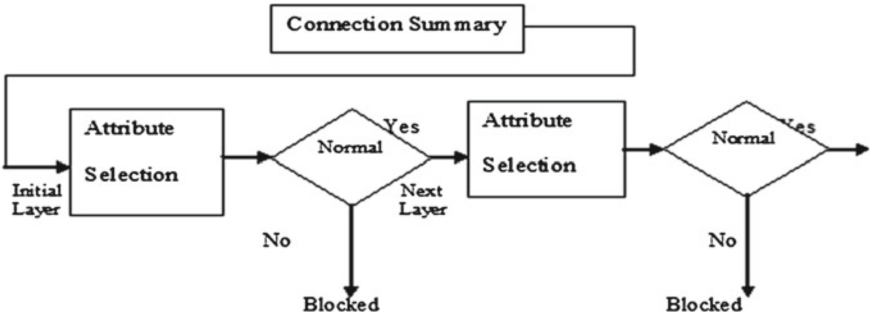


Fig. 2 Layered approach for intrusion detection

the other in a sequence. Similar to this model, the LIDS represents a sequential Layered Approach and is based on ensuring availability, confidentiality, and integrity of data and (or) services over a network. Figure 2 gives a generic representation of the framework. The goal of using a layered model is to reduce computation and the overall time required to detect anomalous events. The time required to detect an intrusive event can significantly be reduced by eliminating the communication overhead among different layers. Every layer in the LIDS framework is trained separately and then deployed sequentially (Table 1).

We define four layers that correspond to the four attack groups [10]. They are Probe layer, DoS layer, R2L layer, and U2R layer. Each layer is then separately trained with features. Feature selection is significant for Layered Approach. The layers essentially act as filters that block any anomalous connection, thereby eliminating the need of further processing at subsequent layers enabling quick response to intrusion. The effect of such a sequence of layers is that the anomalous events are identified and blocked as soon as they are detected. Hence, we implement the LIDS and select a small set of features for every layer rather than using all the 41 features. This results in significant performance improvement during both the training and the testing of the system. In many situations, there is a trade-off between efficiency and accuracy of the system and there can be various avenues to improve system performance. Methods such as naive Bayes [9] assume independence among the observed data. This certainly increases system efficiency, but it may severely affect the accuracy. To balance this trade-off, we use the CRFs that are more accurate, though expensive, but we implement the Layered Approach to improve overall system performance. The performance of our proposed system, Layered CRFs is comparable to that of the decision trees and the naive Bayes, and our system has higher attack detection accuracy.

3.2 Conditional Random Fields for Intrusion Detection

Conditional models are systems that are used to model the conditional distribution over a set of random variables. Such models have been extensively used in the natural language processing tasks. Conditional models offer a better framework as they do not make any unwarranted assumptions on the observations and can be used to model rich overlapping features among the visible observations. CRFs are undirected graphical models used for sequence tagging. The CRFs have proven to be very successful in such tasks, as they do not make any unwarranted assumptions about the data. Hence, we explore the suitability of CRFs for intrusion detection. System may consider features such as "logged in" and "number of file creations." When these features are analyzed individually, they do not provide any information that can aid in detecting attacks. However, when these features are analyzed together, they can provide meaningful information.

3.3 Integrating Layered Approach with Conditional Random Fields

A natural choice is to integrate them to build a single system that is accurate in detecting attacks and efficient in operation.

Probe Layer

The probe attacks are aimed at acquiring information about the target network from a source that is often external. Hence, basic connection level features such as the "duration of connection" and "source bytes" are significant while features like "number of files creations" and "number of files accessed" are not expected to provide information for detecting probes (Fig. 3).

DoS Layer

For the DoS layer, traffic features such as the "percentage of connections having same destination host and same service" and packet level features such as the "source bytes" and "percentage of packets with errors" is significant.

R2L Layer

The R2L attacks are one of the most difficult to detect as they involve the network level and the host level features. We therefore select both the network level features such as the "duration of connection" and "service requested" and the host level

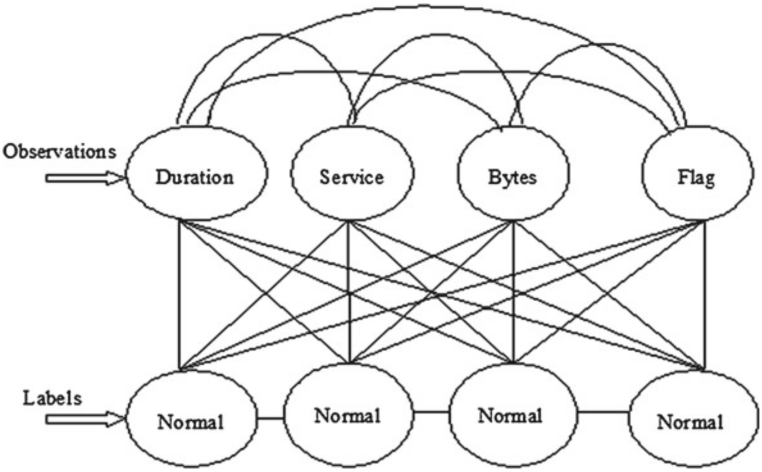


Fig. 3 Conditional random field

features such as the "number of failed login attempts" among others for detecting R2L attack.

U2R Layer (User to Root Attacks)

The U2R attacks involve the semantic details that are very difficult to capture at an early stage. Such attacks are often content based and target an application. Hence, for U2R attacks, we select features such as "number of file creations" and "number of shell prompts invoked," while we ignored features such as "protocol" and "source bytes."

3.4 Intrusion Detected Message Sent to System Administrators Mobile

The mobile device can be used to keep oneself informed about the attacks. The corresponding error messages are generated and are intimated to the server which schedules the appropriate actions. Mobile alerts are sent to the server administrator's mobile through usage of a GSM modem connected to the com port of your computer and making sure that the java communication API is installed in your system. We also carefully consider several parameters such as text message centre number found in your mobile in the SMS settings menu and the baud rate and Type of flow control for receiving, Type of flow control for sending, the number of data bits, the number of stop bits, and the type of parity (Fig. 4).

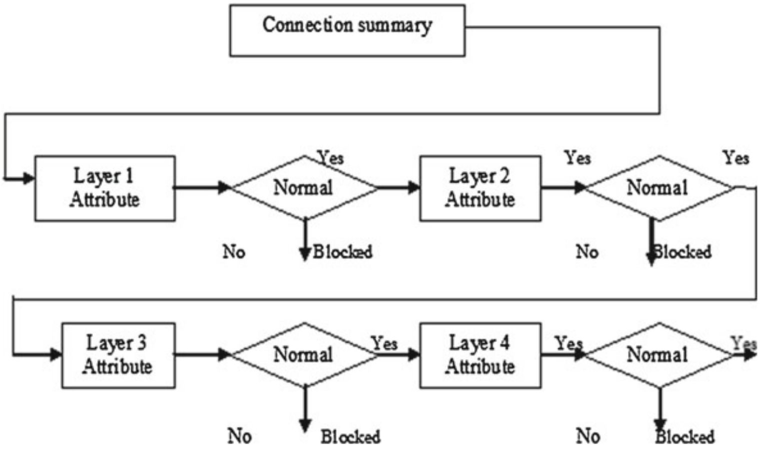


Fig. 4 Integrating layered approach with conditional random fields

3.5 *Proposed Algorithm*

Step 1: Select the number of layers, n, for the complete system.
Step 2: Separately perform features selection for each layer.
Step 3: Plug in the layers sequentially such that only the connections labelled as normal are passed to the next layer.
Step 4: For each (next) test instance perform Steps 5 through 8.
Step 5: Test the instance and label it either as attack or normal.
Step 6: If the instance is labelled as attack, block it and identify it as an attack represented by the layer name at which it is detected and go to Step 4. Pass sequence to next layer.
Step 7: If the current layer is not the last layer in the system, test the instance and go to Step 6. Else go to Step 8.
Step 8: Test the instance and label it either as normal or as an attack. If the instance is labelled as an attack, block it and identify it as an attack corresponding to the layer name.
Step 9: If the instance is labelled as an attack at any layer, then intimate it to system admin's mobile with a corresponding appropriate message of attack.

4 Results

From the above test suite it can be concluded that our proposed system is capable of detecting intrusions at the various layers by using layered conditional random fields and when detected they will be first intimated to the system administrator at

Table 1 Test Case Results

Test case	Input	Action	Expected result	Obtained result
(1) New user creation	(1) Enter the user details as a required as follows: Name, Age, Gender, Address, Phone No., Access time, Password	(1) Click OK	User details must be added to database	User details are added to database
		(2) Click cancel	New user page would close	New user page closed
		(1) Click OK	User details must not be added to database	User details are not added to database
	(2) User details not entered correctly	(2) Click cancel	New user page would close	New user page closed
(2) Login page working	(1) Enter user ID, Password	(1) Submit	Profile page must be displayed	Profile page displayed
		(2) Reset	Login page must be displayed	Login page is reset
		(3) Exit	Login page would close down	Login page closed
	(2) Password of user ID not entered	(1) Submit	Enter the key alert message must be displayed	Alert messages displayed
		(2) Reset	Login page must be reset	Login page is reset
		(3) Exit	Login page would close	Login page closed
(3) Profile page working	(1) P1-if authenticated	Click—get result	System properties must be displayed	System properties displayed
	(2) P2-If authenticated	Click—get result	Memory information must be displayed	Memory Information displayed

(Continued)

Table 1 (Continued)

Test case	Input	Action	Expected result	Obtained result
(4) Sever home page working	Status message	(1) Click—clear	Status messages must be cleared	Status messages cleared
		(2) Click—store	Status messages must be stored in the database	Status messages stored in database
(5) Sending a file	(1) Path of a file within the maximum packet range	Click—send	Server must receive the file, process status must display completed message	Server received the file, process status displayed complete message
	(2) If exceeding the maximum packet range	Click—send	Server must get an alert message, process status must display aborted message	Server got an alert message, process status must display aborted message
(6) Login attempt for authorized user	Correct user ID and password	(1) Click submit in login page accordingly, on fourth click	Network is busy alert message must be displayed	Network is busy alert message displayed
		(2) Click back in profile page accordingly, on the fourth click	Network is busy alert message must be displayed	Network is busy alert message displayed
(7) Login attempt for unauthorized user	Wrong user ID and password	Click on submit button in login page accordingly, on the seventh click	Login page must be shut down	Login page shut down

the server side so that necessary actions can be taken. The particular intruder will be denied of access thereby indicating that with the use of mobile phone intimation and layered conditional random fields the system can be more secure.

5 Conclusion

As security breach incidents become more frequent, IDS tools are becoming increasingly necessary. They round out the security factor, working in conjunction with other information security tools, such as firewalls, and allow for the complete supervision of all network activity. In our project we have implemented the test suite using mobile phones for building robust and efficient intrusion detection systems by implementing the layered approach and conditional random fields.

Ideally, the best IDS tools combine both approaches. By this way, the user gets comprehensive coverage, making sure to guard against as many threats as possible. It is clear that using intrusion detection systems is an important and necessary tool in the security manager's arsenal.

Our system addresses the main problem of Efficiency with the best optimal solution, which is not present in the existing systems. The Layered Approach is a signature based system and the Conditional Random Fields is an anomaly based system thus combining these both systems would result in a hybrid system providing both efficiency and accuracy [11, 12]. We also have demonstrated how our system is implemented in real life.

Our system can help in identifying an attack once it is detected at a particular layer, which expedites the intrusion mechanism, thus minimizing the impact of an attack. Once the attack is detected, it is intimated through mobile phone to the system administrator for safe guarding the server system. This type of a system is very much suited in an organizational network. Finally, our system has the advantage that the number of layers can be increased or decreased depending upon the environment in which the system is deployed, giving flexibility to the network administrator.

Acknowledgments The authors sincerely thank the authorities of Supercomputer Education and Research Center, Indian Institute of Science for the encouragement and support.

References

1. Scarfone K, Mell P (2007) Recommendations of the National Institute of Standards and Technology: intrusion detection systems basics. http://csrc.nist.gov/publications/nistpubs/800-94/SP800-94.pdf
2. SANS Institute (2012) Intrusion detection FAQ. http://www.sans.org/resources/idfaq/
3. Gupta KK, Nath B, Kotagiri R (2010) Conditional random fields for intrusion detection. Proceedings of IEEE dependable and secure computing

4. Anderson JP (2010) Computer security threat monitoring and surveillance. http://csrc.nist.gov/publications/history/ande80.pdf
5. Agrawal R, Imielinski T, Swami A (1993) Mining association rules between sets of items in large databases. Proc ACM SIGMOD 22(2):207–216
6. Gupta KK, Nath B, Kotagiri R (2006) Network security framework. Int J Comput Sci Netw Secur 6(7B):151–157
7. Gupta KK (2009) Robust and efficient intrusion detection systems. http://ww2.cs.mu.oz.au/kgupta/files/phd-completion.pdf
8. Gupta KK, Nath B, Kotagiri R (2010) Layered approach using conditional random fields for intrusion detection. Proceedings of IEEE dependable and secure computing
9. Amor NB, Benferhat S, Elouedi Z (2004) Naive Bayes vs. decision trees in intrusion detection systems. In: Proceedings of ACM symposium on applied, computing (SAC'04), pp 420–424
10. Abraham T (2001) IDDM: intrusion detection using data mining techniques. http://www.dsto.defence./gov.au/publications/2345/DSTO-GD-0286.pdf
11. Gupta KK, Nath B, Kotagiri R (2010) Layered approach using conditional random fields for intrusion detection. IEEE Trans Depend Secure Comput 1(7):35–49
12. Saravanan C, Shivsankar MV, Tamije Selvy P, Anto S (2012) An optimized feature selection for intrusion detection using layered conditional random fields with MAFS. Int J Mob Netw Commun Telematics 2(3):79–91

An Efficient Microaggregation Method for Protecting Mixed Data

S. K. Chettri and B. Borah

Abstract *MDAV2k* is an efficient multivariate data-oriented microaggregation technique for microdata protection. However, the method works only on numeric data thus prohibiting it from being used to protect real world data containing mixed values. In this paper we introduce an improved data-oriented microaggregation method for protecting microdata in the case that the records under consideration are expressed in terms of continuous and categorical attribute values. Experimental results shows that the *MDAV2k* method with improved mixed distance measurement can microaggregate mixed data to achieve better *k*-anonymity with less information loss and a better trade-off between information loss and data disclosure risk.

1 Introduction

Protecting the privacy of individuals is a persistent issue for the last many years as more and more people are concerned with privacy issues. At the same time, there is an increasing demand of data by researchers and decision makers for strategic planning and decision making in various fields like in areas of health, marketing, official statistics and so on. The databases may be released for various analysis but the privacy of individuals may be breached if the databases related to them get disseminated, as the information contained in the databases may be sensitive. As shown in [1] there

S. K. Chettri (✉)
Department of Computer Science, Saint Mary's College, Shillong, India
e-mail: s.chettri@smcs.ac.in

B. Borah
Department of Computer Science and Engineering, Tezpur University, Tezpur, India
e-mail: bgb@tezu.ernet.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_54,

exist several kinds of disclosure risks to the privacy of the individuals. Thus the problem is, how to extract relevant knowledge from large amount of data while protecting at the same time sensitive information existing in the database, which may raise various social and ethical issues if got revealed. A popular and efficient approach for preserving individual privacy and resisting linked attack is k-anonymity, which was introduced by Samarati [17] and Sweeney [19]. A released record is said to adhere to k-anonymity if the values of each released record is indistinct from at least $(k-1)$ other records over a special set of fields called quasi-identifier. In this context a new method called microaggregation has been proposed in the literature. Microaggregation is a family of Statistical Disclosure Control (SDC) [8] method which naturally satisfies k-anonymity. It is a perturbative technique which partitions a given dataset into small groups, each consisting of at least k records, with k being a user-defined parameter. The microaggregated dataset is built by replacing each original record by the centroid of the group to which it belongs to. The microaggregated dataset can be released without jeopardizing the privacy of the individuals which form the original dataset because k records have an identical protected value. To minimize information loss caused by microaggregation as it involves data modification or masking, records within each group should be as homogeneous as possible. For being a perturbative method the challenge in microaggregation is to tune the modification of data so that both privacy risk and information loss are kept below certain acceptable limits.

Multivariate microaggregation with maximum within group's record homogeneity is NP-hard [15], so heuristics are normally used. There exist two main types of heuristics: fixed-size microaggregation [2, 5–7, 12, 14] and data-oriented microaggregation [4, 13, 18]. The data-oriented microaggregation method in contrast with their fixed-size counterparts' yield groups of variable sizes depending on the distribution of the original records. It achieves lower information loss by adapting the choice of group sizes to the structure of the dataset. *MDAV2k* [3] is an efficient multivariate data-oriented microaggregation technique for microdata (i.e. individual records and/or company records) protection. However, the method works only on numeric data thus prohibiting it from being used to protect real world data containing mixed values. This paper introduces an improved *MDAV2k* method for protecting microdata in the case that the records under consideration are expressed in terms of continuous and categorical attribute values.

The rest of the paper is organized as follows. Section 2 gives some related concepts and mathematical preliminaries of the algorithm in conjunction with mixed data. In Sect. 3, we discuss the *MDAV2k* algorithm. In Sect. 4, we present some measures to evaluate the microaggregation methods related to numeric, categorical and mixed data. In Sect. 5, experimental data and results are presented and the effectiveness of the proposed modified algorithm is assessed. Finally, in Sect. 6 conclusions are drawn.

2 Microaggregation Techniques

2.1 Related Concepts

Definition 1 (**Quasi-identifier**) A quasi-identifier (QI) is a set of attributes in a dataset D such that the set can be used to identify individual records in D by linking with an external information. For e.g. job, age, gender etc.

Definition 2 (***k*-anonymity**) A dataset D is said to satisfy k-anonymity for $k > 1$ if, for each combination of values of quasi-identifiers (QI) there exist at least k records in the dataset sharing the same combination thus making each record indistinguishable from at least $(k - 1)$ other records.

k-anonymity can be achieved through generalization or suppression. In generalization a value is replaced or recoded with a less specific but semantically consistent value while in suppression a value is not released at all. Thus microaggregation is found to be an efficient perturbative SDC method which naturally satisfies k-anonymity. The basic concept of microaggregation can be defined in terms of the following two steps

- Partition: Given a dataset D with n individual records, it is partitioned into several clusters with each cluster consisting at least k records.
- Aggregation: Records in each cluster is replaced with the aggregate value of the cluster. It has been proved in [6] that optimal partition yields cluster of size between k and $2k - 1$.

The main goal of microaggregation method is to produce a microaggregated dataset D' from original dataset D in such a way that the data is protected from being disclosed to an intruder and at same time the user analyses on D and D' yields same or similar results.

2.2 Distance Measurement

2.2.1 Objects with Numerical Attributes

Let X, Y be two tuples representing a total number of n numerical values. The distance between X, Y is the Euclidean distance as

$$d_n(X, Y) = \sum_{i=1}^{n} \sqrt{(x_i - y_i)}^2 \tag{1}$$

The centroid Z_i of the cluster C_i is computed by finding the mean of the numerical attribute values in the cluster as

$$Z_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij} \tag{2}$$

where n_i is the number of numerical attributes in cluster C_i and x_{ij} is the jth attribute of ith cluster.

2.2.2 Objects with Categorical Attributes

Let X, Y be two tuples representing a total number of p categorical values. The simple matching distance measure between X, Y as given in [7] is

$$d_c(X, Y) = \sum_{j=1}^{p} \delta(x_j, y_j) \tag{3}$$

where $\delta(x_j, y_j) = \begin{cases} 0 \ (x_j = y_j) \\ 1 \ (x_j \neq y_j) \end{cases}$

Object Allocation to Cluster

If simple matching distance for categorical attributes as defined in (3) is used to find the nearest cluster of an object to be assigned, then it will lead to weak intra-cluster similarity as shown in the Example 1 below

Example 1 Let C_1 and C_2 be two clusters with three categorical attributes A_1, A_2 and A_3

$$C_1\text{: mode [a, c, e]} \begin{bmatrix} A_1 & A_2 & A_3 \\ a & c & e \\ a & c & f \\ b & d & g \end{bmatrix} \quad C_2\text{: mode [a, c, e]} \begin{bmatrix} A_1 & A_2 & A_3 \\ a & c & e \\ a & c & f \\ a & c & g \end{bmatrix}$$

If an object X = [a,c,f] is to be assigned to the available clusters then according to k-modes algorithm using simple matching distance as in (3), the object X can be assigned to any cluster C_1 or C_2, but the intra-cluster similarity of cluster will be more if object X is assigned to C_2.

Thus to preserve the implicit similarity relationship embedded between a categorical tuple X_i and cluster C_l having p categorical attributes as in [10], we compute a score s_w by attribute weighting function between the tuple X_i and cluster C_l as

$$s_w(X_i, C_l) = \sum_{j=1}^{p}(1 - \omega(x_{ij}, l) \tag{4}$$

where $\omega(x_{ij}, l)$ is a weight value of $x_{i,j}$ in cluster C_l, x_{ij} represents jth attribute of tuple X_i, weight is calculated as relative frequency of the attribute values of tuple X_i in cluster C_l as $\omega(x_{ij}, l) = f(x_{ij}|C_l)/|C_l|$.

The centroid selection Z_l of cluster C_l containing categorical attributes is normally done through computing frequency mode of attribute values in each cluster as $f(a_j^r|C_l)$ where $a_j \in A_j$ and A_j is a set of categorical attribute values denoted by a domain Dom $(A_j) = \{a_j^{(1)}, a_j^{(2)}, a_j^{(3)} \ldots\ldots a_j^{(p_j)}\}$, p_j is the number of categorical values of A_j.

2.2.3 Objects with Mixed Attributes

Let $X = \{x_1, x_2, x_3 \ldots x_n, x_{n+1}, x_{n+2} \ldots x_p\}$ and $Y = \{y_1, y_2, y_3 \ldots y_n, y_{n+1}, y_{n+2} \ldots y_p\}$ be two tuples representing a total number of n numerical values and p categorical values, where $\{x_1, x_2, x_3 \ldots x_n\}$ and $\{y_1, y_2, y_3 \ldots y_n\}$ represent numerical attributes of X and Y respectively. Similarly, $\{x_{n+1}, x_{n+2} \ldots x_p\}$ and $\{y_{n+1}, y_{n+2} \ldots y_p\}$ represent categorical attributes of X and Y respectively. The distance is calculated as

$$d_m(X, Y) = d_n(x_{1\ldots n}, y_{1\ldots n}) + \gamma_b d_c(x_{n+1\ldots p}, y_{n+1\ldots p}) \tag{5}$$

where d_n represents Euclidean distance of numerical attributes as defined in (1) and d_c represents simple matching distance measure of categorical attribute values as defined in (3), γ_b is a coefficient to balance the proportion of numerical and categorical attribute values of tuples X and Y.

The balance coefficient γ_b has a role to play in clustering mixed data objects and its behavior has been studied in [11]. If $\gamma_b = 0$, clustering depends only on the numerical attributes. Here, we have taken γ_b depending on the normal distributions of the numeric attributes in the dataset. It is calculated as the average standard deviation σ of the numeric attributes in dataset D.

Object Allocation to Cluster

A tuple $X = \{x_1, x_2, x_3 \ldots x_n, x_{n+1}, x_{n+2} \ldots x_p\}$ is allocated to a cluster C_m with centroid $Z_m = \{z_1, z_2, z_3 \ldots z_n, z_{n+1}, z_{n+2} \ldots z_p\}$ if it has a minimum distance to it. The distance is calculated as

$$d_{nw}(X, C_m) = d_n(x_{1\ldots n}, z_{1\ldots n}) + \gamma_b s_w(X, C_m) \tag{6}$$

where d_n represents Euclidean distance of numerical attributes as defined in (1) and s_w represents a score by attribute weighting function of categorical attribute values of tuple X and cluster C_m as defined in (4). And the balancing coefficient γ_b is calculated as the average standard deviation σ of the numeric attributes.

The Centroid Z_m of the cluster C_m is computed as mean of the numerical attributes in cluster C_m as defined in (2) combined with the selection of categorical attributes of cluster based on frequency mode of attribute values in each cluster as shown above.

3 *MDAV2k* Algorithm for Mixed Data

*MDAV*2*k* as seen in [3] is an efficient multivariate data-oriented microaggregation method. Being a data-oriented method it partitions the dataset into groups of variable sizes depending on the distribution of the original records where γ value is dynamically computed to conservatively expand the group. The complexity of the method has been proved to be $O(n^2)$. We have improved the *MDAV2k* method to protect mixed data. The algorithm is given as under

Algorithm 0.1 *MDAV2k*

```
1. Compute centroid $\bar{x}$ of records in dataset $D$. Find the most distant record $x_r$ from $\bar{x}$.
2. Find $2k$ nearest records $(y_1, y_2 ... y_{2k})$ of $x_r$ and form a group $c_i$ around $x_r$ with its $k-1$ closest records.
3. Compute centroid $\bar{x}_i$ of $c_i$. Find distance $d_1$ from $\bar{x}_i$ to $x_r$ and distance $d_2$ from $x_i$ to $y_j$ where$j = k$.
4. Find $k-1$ nearest records $(z_1, z_2 ... z_{k-1})$ of $y_j$ and compute a centroid $\bar{z}$ of the temporary formed grouped around $y_j$.
5. Find distance $d_3$ from $\bar{z}$ to $y_j$. Compute $\gamma = d_3/d_1$ and restrict the value of $\gamma$ within 1.16 by setting $\gamma = 1 + 1/(5+\gamma)$ if $\gamma > 1$.
6. Check if $(d_2 < \gamma d_3)$ then record $y_j$ is inserted into group $c_i$.
7. If the group $c_i$ contains less than $2k$ records than go to step 3 taking next $k+1$ nearest record of $x_r$.
8. If there are atleast $3k$ records remained to form any group then repeat the steps through step 1 taking a new set of records after removing the formed groups.
9. If there remained atleast $2k$ records which do not belong to any group then compute centroid $\bar{x}$ of remaining records in dataset $D$. Find the most distant record $x_r$ from $\bar{x}$. Form a group around $x_r$ with its $k-1$ nearest neighbours.
10. If there exist atleast $2k$ records which do not belong to any group then form a group with the remaining records.
```

4 Evaluating Microaggregation Method

4.1 Information Loss Measurement for Numerical Data

In [6], optimal microaggregation is defined as the one yielding a k-partition maximizing the within-groups homogeneity; the higher the within-groups homogeneity, the lower the information loss, since microaggregation replaces values in a group by the group centroid. They showed that groups in an optimal k-partition contain between k and $2k-1$ records. The sum of square error (*SSE*) criterion is common to measure homogeneity in clustering as shown in [5]. In terms of *SSE*, maximizing within-groups homogeneity is equivalent to finding a k-partition minimizing the within group *SSE*. The goal of microaggregation is to minimize the *SSE* measure, which is defined as

$$SSE = \sum_{i=1}^{g} \sum_{x_{ij} \in C_i} (x_{ij} - \overline{x}_i)'(x_{ij} - \overline{x}_i) \tag{7}$$

where g is the total number of number of clusters, C_i is the ith cluster and $\overline{x}_i$ is the centroid of C_i.

The total sum of square *SST* is the sum of square error within the entire dataset calculated by summing the Euclidean distance of each record x_{ij} to the centroid $\overline{x}$ of dataset D is as follows

$$SST = \sum_{i=1}^{g} \sum_{x_{ij} \in C_i} (x_{ij} - \overline{x})'(x_{ij} - \overline{x}) \tag{8}$$

The value *SST* for a dataset does not changes but *SSE* changes with partitioning dataset into different groups. Microaggregation techniques are often compared on the basis of the *SSE* or the IL_n (Information Loss) measure. The measure IL_n standardized between 0 and 100 can be obtained as

$$IL_n = \frac{SSE}{SST} \cdot 100 \tag{9}$$

4.2 Information Loss Measurement for Categorical Data

The IL_c (Information Loss) measure of categorical data can be defined as

$$IL_c = \frac{1}{g} \sum_{i=1}^{g} \frac{d_c(X_i, X_i')}{|C_i|} \cdot 100 \tag{10}$$

where g is the total number of number of clusters, X_i denotes the ith categorical tuple of cluster C_i and X_i' is the anonymized version of original tuple X_i, $|C_i|$ is size of cluster C_i.

4.3 Information Loss Measurement for Mixed Data

Average IL_m (Information Loss) measurement for mixed data can be defined as

$$IL_m = \frac{IL_n + IL_c}{2} \tag{11}$$

4.4 Data Disclosure Risk Measurement

Data disclosure risk measurement is used to assess the security of anonymized table. We adopt here the Distance Linkage Disclosure Risk (DLD) model as in [16]. It is based on the probability of inferring the original record from the anonymized table. It can be defined as for any anonymized record X' in an anonymized table D' if we compute a distance to other records in the original table D, we can get a nearest record X_1 and a second nearest record X_2. If X_1 or X_2 is the original record X, then the record X is called a linked_record.

Let *num_linked_record* be the number of linked records in an anonymity table, *total_num_record* be the total number of records in an anonymity table, then *DLD* is defined as

$$DLD = \frac{num\_linked\_record}{total\_num\_record} \cdot 100 \tag{12}$$

5 Data and Results

We have implemented the modified *MDAV2k* algorithm in C under Linux environment on a machine with 2.13 GHz Intel i3 Processor and 3 GB RAM. We have used the publicly available dataset, Adult Dataset, [9] from the UCI Machine Learning Repository. The dataset contains 48,842 instances with 14 attributes from US Census data, which is a commonly used dataset for testing clustering and classification algorithms.

After cleaning the dataset of missing values, we have got 45,222 instances with 10 attributes which are standardized by the z-score method where the mean μ is subtracted from individual attribute values and the result is then divided by the standard deviation σ of the population. This is done to ensure that all attributes have equal weights when computing distances. Table 1 gives an insight of the attributes

Table 1 Adult dataset description

No	Attribute	Data type	Distinct values
1	Age	Continuous	74
2	Fhlweigh	Continuous	100
3	Education_num	Continuous	16
4	Hours_per_week	Continuous	99
5	Work class	Categorical	8
6	Education	Categorical	16
7	Marital status	Categorical	7
8	Occupation	Categorical	14
9	Race	Categorical	5
10	Sex	Categorical	2

Table 2 Experimental results with different k values

Dataset	Measures	$k = 3$	$k = 5$	$k = 10$	$k = 15$	$k = 20$
	SSE	789.484	1167.428	1597.563	1821.258	1930.949
	IL_n	5.456	8.068	11.04	12.586	13.344
"Adult"	IL_c	87.023	87.146	87.359	87.714	87.927
	IL_m	46.239	47.607	49.2	50.150	50.635
	DLD	6.479	3.288	1.46	0.805	0.584

we have used from the dataset. The "distinct values" column indicates the different values which an attribute can have.

The algorithm was analyzed based on cluster homogeneity SSE, information loss measures - IL_n, IL_c, IL_m and disclosure risk DLD of the microaggregated dataset with different values of k. The results are presented in Table 2.

5.1 *Information Loss and Data Disclosure Risk for Different* k *Values*

In order to compare the performance of the algorithms both Information Loss (IL) and Distance Linkage Disclosure risk (DLD) measures are reported for different values of k as seen in Figs. 1 and 2.

Figure 1 shows the performance curve of average Information Loss *(IL)* over different k values. The average information loss increases with the increasing k value. Figure 2 shows the performance curve of data disclosure risk over different values of k which clearly indicates that the data disclosure risk decreases with the increasing k value.

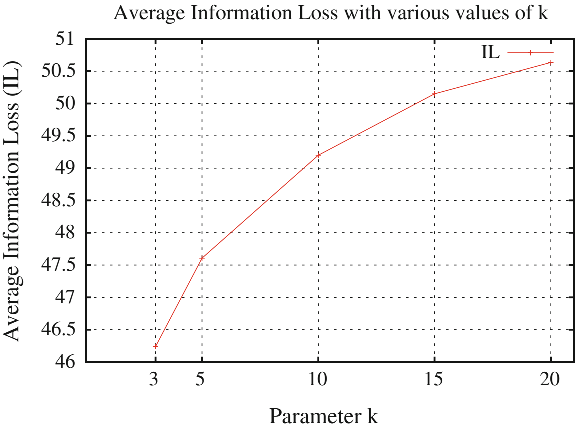


Fig. 1 Average information loss for different k values

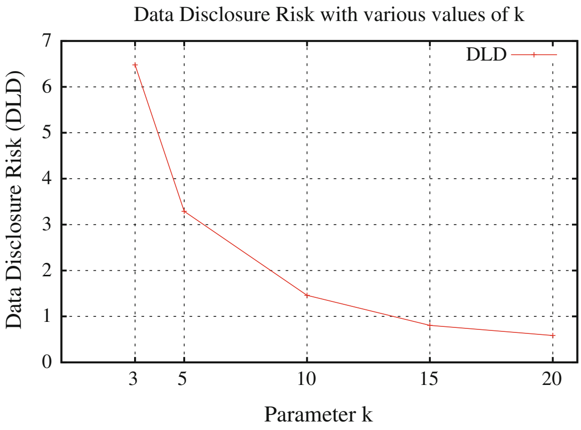


Fig. 2 Data disclosure risk for different *k* values

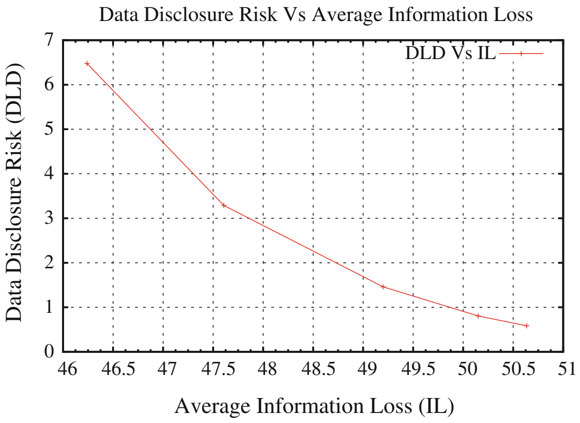


Fig. 3 Data disclosure risk versus information loss

5.2 *Data Disclosure Risk Versus Information Loss*

We have also compared the algorithm based on its data disclosure risk and information loss to show a trade-off between the two measures. Figure 3 shows the performance curve of the algorithm based on the trade-off. It can be seen that the data disclosure risk decreases with increasing information loss.

6 Conclusion

In this paper we have proposed an improved *MDAV2k* method which is a multivariate data-oriented microaggregation technique to protect mixed data. It investigates the distance measurement of mixed data and focuses mainly on improving the intra-cluster similarity when categorical data are involved in clustering. We have made the experimental analysis with a publicly available standard dataset. The experimental results shows that the improved *MDAV2k* method can efficiently protect mixed data by achieving a better trade-off between information loss and data disclosure risk.

References

1. Aggarwal CC, Pei J, Zhang B (2006) On privacy preservation against adversarial data mining. In: Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining–KDD '06 . ACM Press, New York, p 510
2. Chang CC, Li YC, Huang WH (2007) TFRP: an efficient microaggregation algorithm for statistical disclosure control. J Syst Softw 80(11):1866–1878
3. Chettri SK, Borah B (2012) MDAV2K: a variable-size microaggregation technique for privacy preservation. International conference on information technology convergence and services, In, pp 105–118
4. Domingo-Ferrer J (2006) Privacy in statistical databases: k-anonymity through microaggregation. IEEE granular computing, Atlanta, In, pp 747–777
5. Domingo-Ferrer J, Martinez-Balleste A, Mateo-sanz JM, Sebé F (2006) Efficient multivariate data-oriented microaggregation. VLDB J 15:355–369
6. Domingo-Ferrer J, Mateo-Sanz JM (2002) Practical data-oriented microaggregation for statistical disclosure control. IEEE Trans Knowl Data Eng 14(1):189–201
7. Domingo-Ferrer J, Torra V (2005) Ordinal, continuous and heterogeneous k-anonymity through microaggregation. Data Min Knowl Discov 11(2):195–212
8. Fayyoumi E (2010) A survey on statistical disclosure control and microaggregation techniques for secure statistical databases. Softw Pract Exper 40:1161–1188
9. Frank A, Asuncion A (2010) UCI machine learning repository
10. He Z, Xu X, Deng S (2011) Attribute value weighting in k-modes clustering. Expert Syst Appl 38(12):15365–15369
11. Huang Z (1997) Clustering large data sets with mixed numeric and categorical values. Asia conference on knowledge discovery and data, In, pp 1–14
12. Laszlo M (2005) Minimum spanning tree partitioning algorithm for microaggregation. Knowl Data Eng 17:902–911
13. Lin J, Wen T (2010) Density-based microaggregation for statistical disclosure control. Expert Syst Appl 37(4):3256–3263
14. Nin J (2008) On the disclosure risk of multivariate microaggregation. Data Knowl Eng 64:346–364
15. Oganian A, Domingo-Ferrer J (2001) On the complexity of optimal microaggregation for statistical disclosure control. Stat J U N Econ Comm Eur 18(4):345–354
16. Pagliuca D (1999) Some results of individual ranking method on the system of enterprise accounts annual survey. Esprit SDC Project, Deliverable MI-3/ D
17. Samarati P (2001) Protecting respondents identities in microdata release. IEEE Trans Knowl Data Eng 13:1010–1027
18. Solanas A (2006) V-MDAV: a multivariate microaggregation with variable group size. Seventh COMPSTAT symposium of the IASC. Springer, Rome
19. Sweeney L (2002) k-anonymity: a model for protecting privacy. Int J Uncertain Fuzziness Knowl Syst 10(5):1–14

Plus/Delta (+/Δ) Evaluation to Help Organizations Deliver Projects Effectively

A. Pathanjali Sastri and K. Nageswara Rao

Abstract One of the most serious complaints against software failure is the inability to estimate with acceptable accuracy the cost, resources, and schedule necessary for a software project [1]. The less visible one is the underwater part representing poor quality resulting in significant rework, cost overrun due to high rework and idle times, and customer dissatisfaction [2]. IT projects fail for various reasons and whatever the reason, the project manager has to face a frustrated customer, a displeased senior management, and a project team who think they just wasted a lot of time. The objective of this research paper is to develop an investigation model by building a rich picture of how different project stakeholders perceive the quality of software engineering processes, identify the methods and techniques for supporting the job of Software Engineers and Project Managers and propose improvements to these methods that have a wider view of process quality.

Keywords Action research · Plus-Delta · Project management · Project failures

1 Introduction

In order to make a mark in the area of software engineering, we observe that most of the research is to identify to invent a new model or methodology. But at present we have better models in software engineering and if we critically examine the

A. Pathanjali Sastri (✉)
Department of Computer Application, V.R.Siddhartha Engineering College, Kanuru,
Vijayawada, Andhra Pradesh 520 007, India
e-mail: akellapatanjali@yahoo.com

K. Nageswara Rao
Department of Computer Science and Engineering, P.V.P.Siddhartha Institute of Technology,
Kanuru, Vijayawada, Andhra Pradesh 520 007, India
e-mail: drknrao@ieee.org

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_55,

situation we understand that the problem is inadequate models but we need to apply the best practices consistently and effectively instead. Gerald Weinberg said, “When the thinking changes, the organization changes, and vice versa”. By following an “Action Research” or “Participatory Action Research” which is a recognized form of experimental research that focuses on the effects of the researcher’s direct actions of practice within a participatory community with the goal of improving the performance quality of the community or an area of concern, we conducted a series of interviews with software employees at various levels to understand how different project stakeholders perceive the quality of software engineering processes.

2 Justification of Selecting Action Research and Plus/Delta as a Right Tools for Research

“Action research or participatory action research is a research initiated to solve an immediate problem or a reflective process of progressive problem solving led by individuals working with others in teams or as part of a community of practice to improve the way they address issues and solve problems” [3]. Perhaps the most important aspect of action research is that the process enhances practitioners’ professional development through the development of their capability as professional knowledge makers, rather than simply as professional knowledge users. This will bring about the development in their practices by analyzing the existing practice and identifying elements for change. The process will gather the evidences on which to make informed rather than intuitive judgments and decisions and help project management feel in control of their own professional situation. We then selected Plus/Delta which is a simple, formative evaluation process that is quick, easy to use, and provides ideas for improvements [4].

3 Plus/Delta (+/Δ) Evaluation

The purpose of the process is to make us better. But how will it make us better? By monitoring things we are doing, or analyzing and implementing the things we could do better. These findings help the projects to focus on continuous improvements. Plus/Delta (+/Δ) evaluation is an independent assessment of any designated process or activity and is used to identify value added, non-value added, and non-value added but required activities in the processes.

Plus/Delta (+/Δ) evaluation quality tool is a scientific approach that provides a method for continuous improvement by continuously seeking ways to provide the highest quality services. This method is usually determined through the expectations of the customer, the available resources and circumstances.

3.1 Common Causes of Failure

The Every Company sets up a Quality group to ensure that the project follows the defined processes, and conducts a process compliance audit to highlight areas that are not following the defined process, collect and report metrics data to the organization at the end of every month [5–15]. On one side we have a great deal of shared learning and wisdom available on the process front like adhering to comprehensive quality management system and quality assurance processes, leveraging project experiences and learning to bring about predictability in processes and continuity and sustainability in quality. On the other side there have been several studies into software project failure that have attributed failure to one or more areas of project management. Poor deliverables, schedule slippages, cost overruns, etc. which are the attributes of the failure projects may all be result of process inefficiencies as shown in Table 1. Let's apply the process approach to your question. People don't set out to develop defective software so how do the best efforts of capable and well intentioned people allow a defective product to be released, deployed and operated. So, we followed a +/Δ (Plus/Delta) evaluation quality tool provides a method for examining the process from end to end and find solutions that can be easily implemented to eliminate the root causes of process issues.

Table 1 Common causes of software project failure

S. No	Common failures
1	Poor definition of project scope and objectives; Unclear or Poor incomplete or changing requirements; Scope Creep
2	Insufficient time or funds given to project
3	Insufficient and/or over-optimistic planning; Poor estimating; Long or unrealistic timescales; forcing project end dates despite best estimates
4	Lack of user involvement; Project team members lack experience and do not have the required skills
5	Inappropriate technology choices; lack of required technical skills
6	Integration problems during implementation
7	Poor or insufficient testing before go-live
8	Lack of Quality Assurance for key deliverables
9	Insufficient attention to stakeholders and their needs; failure to manage expectations; lack understanding of the project and not actively involved
10	Inadequate visibility of project status; Inadequate tracking and reporting; not reviewing progress regularly
11	Poor collaboration, communication and teamwork
12	Poor project management best practices; inadequately trained or inexperienced project managers

4 Problem Resolution Investigation Using Delta (+/Δ) Evaluation

We use process models to apply engineering rigor to the project and ensure important activities during the execution of the project. But at the same time an efficient process will not be effective if it does not meet the expectations of both the customer and the management. No doubt we are committed to maintaining the highest quality standards in terms of people, processes, and also infrastructure. Throughout the world every project manager would say that it is his/her constant endeavour to ensure the highest quality in every aspect of IT services that are delivered. Despite of achieving certification of various standards, such as ISO, SEI-CMMI, TL9000, AS9100, etc to showcase themselves that they follow global standards and are committed to exceeding client expectations by continuously seeking ways to provide the highest quality services, the project managers face several vulnerabilities and uncertainties during project execution. So the need of the hour is to have an innovative solution or follow a planned and systematic approach to the evaluation of quality and adherence to software product standards, processes, and procedures. Pluses (+): Identify the processes that are working first for the projects in most of the organizations. In the ideal environment, the following are the processes involved in building any of the products and it can be what projects feel they are performing or helps them/beneficial in executing the projects (+).

a. Establishment of policies, standards, processes, and procedures.
b. Conduct periodic internal audits, reviews, acceptance tests, and
c. Metrics collection and analysis
d. Continuous Improvement
e. Induction of Best Practices
f. Status reports of the project leaders/project managers and Quality representative to management.

Deltas (Δ): What they need (Activities) to change for the process/activities to improve for them (opportunities for improvement). The following steps are followed

a. Scope the Improvement
b. Focus on providing value to the projects
c. Develop an Action Plan
d. Should be action oriented
e. Should be reviewed and acted upon as soon as possible

We examined the reasons as to why software projects go out of control in terms of budget, schedule or effort as discussed in Table 1, and also through the discussions with few project managers to understand their organizations engineering and management practices, we generated a focused list of the critical areas for improvement i.e. the things that can be improved/changed (Δ) so that project management may be more effective.

4.1 Practical Solutions for the Software Development Challenges (Deltas-Δ)

A. Understand and negotiate project trade-offs, and track progress: As the user's expectations change periodically over the course of the project, it is extremely important that the customer and the project manager come to an understanding about the assumptions and expectations of the project. The changing project parameters (improve quality, reduce costs meet a particular deadline, or control the scope creep) must be reviewed time to time to make the project successful. But in reality many projects are mismanaged through shortcuts in the development process, poor management of expectations, premature estimates, cost overruns, schedule slippages, or poor resource management. To identify the alternatives for adjusting the project and appreciate the dynamics of changing project parameters the project managers have to use a tool called Management Expectation Matrix [16]. Management Expectation Matrix will help to communicate and discuss with the client what is driving the project at that moment and ensure you stick to this throughout the project. This also helps to keep "the driver" in mind especially during project management reviews.

B. Determine Risks and Plan to Mitigate: There is a need to appoint a person to address a proactive risk management, ability to identify or mitigate risks and focus on risks for large projects and conduct risk assessment by carrying out a full risk analysis. The regular review of risks will ensure the projects that they are managing them, rather than them managing the projects.

C. Keep Focused on the Goals and Problems: A clear project management monitoring and reviewing process along with senior managers, customers, and core people in the project should use a planned versus Actual details. It allows the entire team to monitor how the projects are progressing with specific tasks, time and money. The same must be linked these milestone reviews by quality representatives and project management reviews to see whether the project is still delivering the original project benefits when reviewing your project. If not, It is required to consider re-scoping or if appropriate abandoning the project rather than wasting valuable time, money, and resources working on something that is not working.

D. The project start-up meeting is a significant event in the life of a project: Project start-up meeting is a very short phase of the project management cycle and but a significant event in the life of a project. It should be considered as the most important meeting of the entire project because it allows the Project Manager to set customer expectations. It will be important for the Project Manager to manage these expectations throughout the project and the meeting should focus to avoid conflicts and reconfirm the potential concerns during the RFP stage. The potential conflicts that might arise out of limited understanding of the project and its operating environment could be avoided to the maximum extent by providing the information to the customer and other project stakeholders.

E. Focusing the organization on the critical issues, planning the improvement and effecting change:

    i. Project Managers may use *5 why problem solving tool* [17–19] to find the fundamental root causes of a given problem and determine corrective actions for those key root causes.
    ii. To maintain customer satisfaction and solve internal and external problems completing a corrective action for a customer, proceed with the *8D analysis* [20]. Customers with problems are often even grateful that the company dealt with these problems in such an efficient manner.

Plus/Delta (+/Δ) evaluation which is a powerful process improvement methodology helps to articulate and mitigate those risks most likely to cause our project to fail. For example, using the requirements provided by the customer and his expectations, a method (development process model, estimation tool, and metrics) will be established. Using Plus/Delta (+/Δ) along with the project managers experience and available repository of best practices and lessons learnt (data need to be captured by the organization through various sources), a method is determined and executed. So the Plus/Delta (+/Δ) is treated as a scientific process, that should be used to ensure the final product developed is according to the users specifications.

5 Conclusion

This paper mainly focuses on providing value to the projects and improving the organization and project processes. Initially we started with iterative cycle of data collection and analysis to build a stronger consensus with regard to the quality characteristics identified. In the process, we tried to investigate software project failures and identify the root causes that have attributed failure to one or more areas of project management through some literature survey. Based on our findings, we have understood that most of the project managers agree that despite of well established processes they need to tackle several issues or challenges. We need to learn from our failures and others failures also, so that the project managers can rebound from those experiences and hopefully prevent the same issues from cropping up again may be in the same situations in the same project or future projects. Using Plus/Delta (+/Δ) the projects can evaluate the effectiveness of the processes deployed and followed has value and these findings contribute to continual improvement of the processes in the projects.

References

1. http://ezinearticles.com/?Causes-Of-Software-Project-Failure&id=453814
2. http://www.thehindu.com/business/Industry/article507513.ece
3. http://en.wikipedia.org/wiki/Action-research
4. http://www.uco.edu/academic-affairs/cqi/files/docs/facilitator-tools/plus-delta.pdf
5. Al-Ahmad W, Al-Fagih K, Khanfar K, Alsamara K, Abuleil S, Abu-Salem H (2009) A taxonomy of an IT project failure: root cause. Int Manag Rev 5(1):93–104
6. Linberg KR (1999) Software developer perceptions about software project failure: a case study. J Syst Softw 49(1999):177–192
7. Nasir MHN, Sahibuddin S (2011) Critical success factors for software projects: a comparative study. Sci Res Essays 6(10):2174–2186
8. Procaccino JD, Verner JM, Overmyer SP, Case study:factors for early prediction of software success and failure. http://www.pages.drexel.edu/jmv23/Procaccino.pdf
9. "Why Projects Fail", NASAs Mars Climate Orbiter Project. www.jiscinfonet.ac.uk/InfoKits/infokit-related.../mars-space-probe
10. http://www.oxbridgewriters.com/essays/project-management/the-aim-of-the-failures-method.php
11. Arvind V, Executive Director, Thinksoft Global Services Ltd, Chennai, http://bit.ly/F4TThinksoft, Business Line. http://www.thehindu.com/business/Industry/article507513.ece
12. Who Killed the Virtual Case File? http://www.spectrum.ieee.org/sep05/1455
13. Malone P (2008) http://www.tagonline.org/articles.php?id=259, Project Failure, Jun 9, 2008
14. http://nclarity.net/data/Documentation/pmo.pdf
15. http://www.iag.biz
16. Whitten JL, Bentley LD, Dittman KC (1999) System analysis and design methods, 4th edn. Tata McGraw-Hill Publishing Company Limited, New York, pp 613–617
17. http://www.smh.com.au/news/Breaking/Verdict-on-a-64m-project-failure/2005/04/15/1113509904098.html
18. http://www.mindtools.com/pages/article/newTMC-5W.htm
19. http://www.manufacturing.net/articles/2006/07/lean-tool-box-how-to-five-whys-problem-solving?menuid=282
20. https://www.hawkerbeechcraft.com/supply-chain/files/8D-Training-9-8-2008.pdf

A Secure Routing Protocol for MANETs Against Byzantine Attacks

Gagan Singla and Pallavi Kaliyar

Abstract MANETs are collection of mobile hosts which are self-configurable. Nodes in MANET communicate with each other through wireless channels with no centralized control. MANET nodes rely on multi-hop communication. Security in MANETs is always a very big issue. There are many attacks in MANET due to which the legitimacy of a network is compromised. Many approaches already existed to address these security issues. In this paper an algorithm was proposed for data transmission by using RSA algorithm for authentication purpose, and a blacklist to prevent sending data packets to those nodes which are malicious. Because of this data security is improved. The proposed routing algorithm is more secure as compared to AODV routing algorithm. The performance is analyzed on the basis of various performance metrics like Packets Received and Network Routing Overhead in the network by using the NS2 Simulator.

1 Introduction

A mobile ad-hoc network (MANETs) is an infrastructure less network of mobile nodes in which the nodes communicate with each other by using radio waves. Nodes in MANET [1–4] communicate with each other through wireless channels with no centralized control. With evolution of wireless prevalence in the last decade, we are witnessing more and more applications moving and adapting to wireless methods to communicate. MANET nodes rely on multi-hop communication i.e. uses multiple

G. Singla (✉)
Department of Computer Engineering, Aryabhatta Group of institutions, Barnala, India
e-mail: vipulgog@gmail.com

P. Kaliyar
Department of Computer Engineering, Shri Ram College of Engineering and Technology, Muzaffarnagar, India
e-mail: pallavi.kaliyar@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_56,

nodes within each other's transmission range can communicate directly through radio channels, whereas, those outside the radio range must rely on intermediate nodes to forward messages towards destination. Mobile hosts can move, leave and join the network whenever they want and routes need to be updated frequently due to the dynamic network topology.

In MANETs one of the important issues is routing, i.e. finding a suitable path from source to destination. Due to the rapid growth in use of applications like online gaming, audio/video streaming, VOIP and other multimedia streaming applications in MANETs, it is mandatory to provide required level of security for reliable quality and delivery of data. Providing security in wireless multi-hop networks is much more challenging than in wire line ones mainly due to its dynamic topology, distributed on-the-fly nature, in, multi-hop communication.

There are many attacks in MANET due to which the network can respond in a way we don't want him to. Such as, Black Hole Attacks, Worm Hole Attacks, Byzantine attacks, DoS attacks. Due to these the node is compromised and the network can go down or data loss can occur. In this paper our goal is to make a route discovery algorithm for data transmission which detects the byzantine nodes which are modifying the data, avoids them in transmission path and doesn't let the data packets to get modified or the network to go down. This paper has number of sections. In Sect. 2, the related work done by various researchers has been presented. Proposed algorithm is given in Sect. 3. In Sect. 4, simulation parameters and simulation results has been given. And the conclusion is in Sect. 5.

2 Related Work

In MANETs Security is the main issue and really very important for the data transmission in the network, if the network is not secure data loss of theft can happen. So a mechanism is needed to secure the data. Several researches have worked in this direction and their works is as follows:-

Songbai Lu et al. [5] work and gave an approach named SAODV for the AODV protocol to prevent any security threat in which a node receives the data but then drops it instead of forwarding it by using hash chains. Y. Hu et al. proposed an approach named ARIADNE to deny arbitrary active attackers with the use of symmetric cryptography for the DSR protocol. Sojna Buchegger et al. then proposed another approach namely CONFIDANT for the DSR protocol which could secure the network from the threats which attacks on packet forwarding and routing using the trust based mechanism. Further Ming Yu et al. [6] proposed a scheme for the AODV protocol named as ODSBR which could prevent such attacks those modify the data in the network (Byzantine attacks) with the help of binary search and signatures. After this a better scheme is proposed by S. Yi et al. named as SAR it also prevented modification attacks as well as the replay attacks by using the sequence numbers and trust levels. Then Yih-Chun Hu et al. [7] proposed an approach named

SEAD for the DSDV protocol which prevented the most common Denial of Service attacks using one way hash functions (Table 1).

3 Proposed Algorithm

Here the aim is to provide the security to the data while sending from source to destination. Many other researchers have also worked in this direction, the main advantage of this work over the previous works is that by using this approach the network routing overhead gets decreased. As well as the packet loss in the network is less.

The proposed approach is designed for Ad-hoc On-demand Distance Vector (AODV) routing protocol of MANETs in which the nodes build their trust on each other by comparing the control packets. In this approach at the time of route selection the RSA algorithm output which is unique for each and every packet is put into the header field of Route Request packet (RREQ). At the destination end the same unique output of RSA algorithm is put in to the Route Reply packet (RREP) and then we compare it to decide the packet is secured or corrupted. Due to the uniqueness of the RSA algorithms output it is easy to find that the packet is forwarded by a malicious node or a trust worthy node. For the nodes which are malicious nodes in the network a data is gathered about each node which is known as black list and populate it in the network. Due to this a node does not send any packets to those neighboring nodes which are in a black list.

Proposed algorithm is described as follows:

Step1: Run RSA algorithm and gets its unique output and put it in the header field of RREQ packet.

Step2: Source node S broadcasts a route-request (RREQ) packet to its neighboring nodes.

Step3: The neighboring nodes do the rebroadcasting of RREQ packet with the same unique number in the same field.

Table 1 List of secured protocols

Author name reference	Base/enhanced protocol	Attack	Method used
Songbai Lu et al. [5]	AODV/SAODV	SAODV	Hash chains
Y. Hu et al.	DSR/ARIADNE	ARIADNE	Symmetric cryptography
Sojna Buchegger et al.	DSR/CONFIDANT	CONFIDANT	Trust based
Baruch awerbuch et al. [8]	AODV/ODSBR	ODSBR	Binary search and aggregated signatures
S. Yi et al.	AODV/SAR	SAR	Sequence number, trust level
Yih-Chun Hu et al. [7]	DSDV/SEAD	DoS attack	One-way hash function

Step4: When the RREQ packet reaches to the intended destination, destination puts that unique number in the RREP packet also and sends it back to the source node.
Step5: As the reply is received, comparison of the new fields of RREP and RREQ packets with each other is done.
Step6: If it matches, then the RREP packet is forwarded otherwise that node is added to the blacklist. Next time the node sends route request packets to only those nodes that are not in the blacklist.

4 Simulation and Performance Results

Parameters	Variation of parameters
Area (m × m)	1000 × 1000
Nodes	30, 50, 70
No of malicious node	3, 5
Simulation time (s)	500 s
Node speed (m/s)	7, 10
Pause time (s)	10
Traffic type	CBR
Packet size	512
Protocol used	AODV
Mobility model	Random way point
MAC layer protocol	Mac/802_11
Antenna type	Antenna/omni antenna

4.1 Simulation Results

Packets Received (PR) is the number of packets successfully received by the malicious node. Network Routing Overhead (NRO) is the ratio of the number of routing packets sent for a single data packet in the network.
Figures 1 and 2 here, depicts the number of packets received by the malicious node by varying the speed of all the nodes by 7 and 10 m/s respectively and keeping the number of malicious nodes and keeping the total number of nodes constant.
Results show that the New AODV has less number of packets received as compared to the old AODV no matter what the speed is.
Next Figs. 3 and 4 shows the number of packets received by a malicious node when we have a constant speed and varying number of total nodes i.e. 50 and 70.
Here the results show that PR by malicious nodes in New AODV is less than in the old AODV for whatever is the size of the network. Now rest of the figures shows Network Routing Overheads (NRO) performance for the old and new AODV by taking different scenarios like varying the speed and the total number of nodes in a network.
Figures 5 and 6 shows NRO performance when we vary the mobility speed of nodes in the network.

Fig. 1 Packets received by a malicious node when moving at speed 7

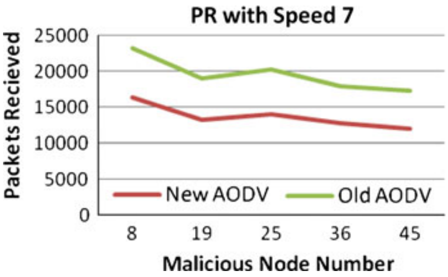


Fig. 2 Packets received by a malicious node when moving at speed 10

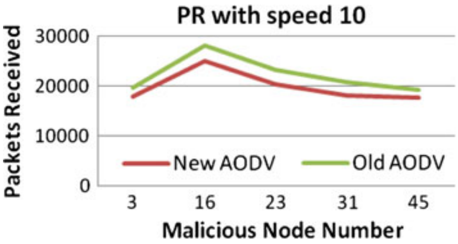


Fig. 3 Showing PR in a 50 nodes network with constant speed

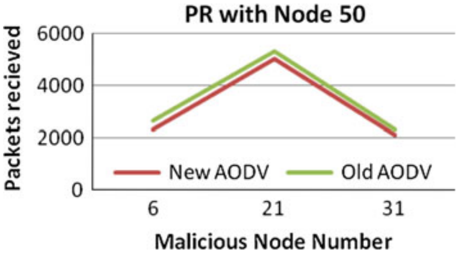
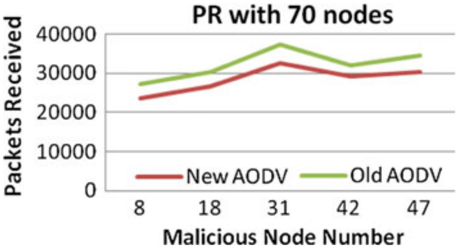


Fig. 4 Shows PR in a 70 node network with constant speed



The results show that the number of Routing Packets sent per data packet is very less in New AODV as compared to old AODV because we do not send any request packet to the nodes which are found as malicious.

Figures 7 and 8 shows NRO when the total number of nodes is increased in a network. Here also the results show that New AODV outperforms Old AODV in all aspects from the very beginning, no matter what's the size of network.

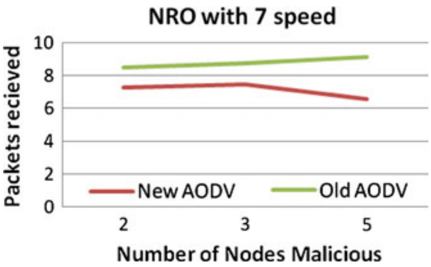


Fig. 5 NRO with speed of nodes 7

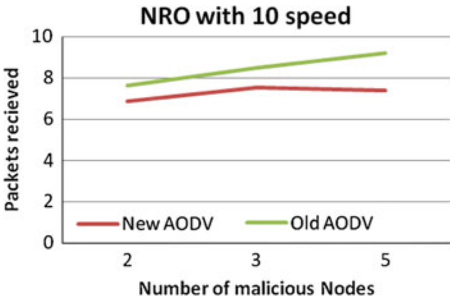


Fig. 6 NRO with speed of nodes 10

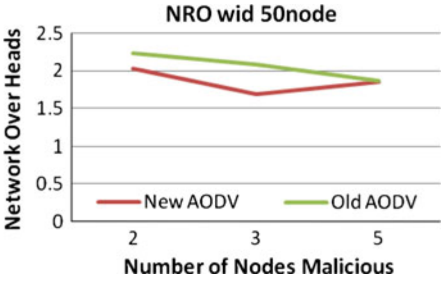


Fig. 7 NRO with total number of nodes increased to 50

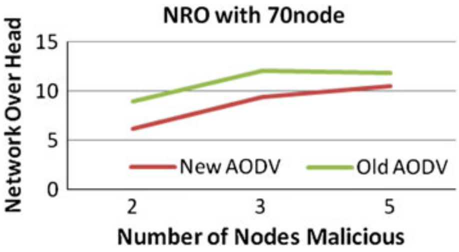


Fig. 8 NRO with total number of nodes increased to 70

5 Conclusion

A lot of work has been already exists in the field of secure data routing over mobile ad-hoc networks [9–20]. Their research work has been considered and an effort has been made to modify the existing AODV protocol by applying RSA algorithm for authentication purpose, and a blacklist to prevent sending data packets to those nodes which are malicious. Because of this data security gets improved. This causes a large decrement in network routing overheads. The proposed algorithm minimizes the packet loss of a node to transfer the data from source to destination, so data security gets increased. The performance of proposed routing protocol is analyzed on the basis of Packet Received and Network Routing Overheads. Simulation results indicate that in the proposed scheme the number of packets received by the malicious nodes is degraded by a significant amount. And that is what is required to do. This Scheme gives best results when CBR traffic is used and the nodes are moving at a high speed. At last, it can be said that by adding the proposed algorithm to the base AODV protocol, a successfully secured AODV protocol is ready up to an extent.

References

1. Taneja S, Kush A (2010) A survey of routing protocols in mobile ad hoc networks. Int J Innov Manag Technol 1(3):279–285
2. Vijaya Kumar G, Vasudeva Reddyr Y (2010) Current research work on routing protocols for MANET: a literature survey. Int J Comput Sci Eng 2(3):706–713
3. Panaousis EA, Ramrekha TA, Millar GP, Politis C (2010) Adaptive and secure routing protocol For emergency mobile ad-hoc networks. Int J Wirel Mobile Netw 2(2):66–78
4. Hu Y-C, Johnson DB, Perrig A (2002) SEAD: secure efficient distance vector routing for mobile wireless ad hoc networks
5. Lu S, Li L, Lam K-Y, Jia L (2009) SAODV: a MANET routing protocol that can withstand black hole attack. International conference on computational intelligence and security
6. Yu M, Zhou M, Su W (2009) A secure routing protocol against byzantine attacks for MANETs in adversarial environments. IEEE Trans Veh Technol 58(1):449–460
7. Awerbuch B, Curtmola R, Holmer D, Nita-Rotaru C, Rubens H (2007) ODSBR: an on-demand secure byzantine resilient routing protocol for wireless ad hoc networks. ACM Trans Inf Sys Secur (TISSEC) 10.4(2008):6
8. Buchegger S, Le Boudec J-Y (2002) Performance analysis of the CONFIDANT protocol: cooperation of nodes-fairness in distributed ad-hoc neTworks. Proceedings of IEEE/ACM workshop on mobile ad hoc networking and computing (MobiHOC), IEEE, Lausanne, CH, June
9. Luo J, Fan M, Ye D (2008) Black hole attack prevention based on authentication mechanism, ICCS 2008 IEEE
10. Jaafar MA, Zukarnain ZA (2009) Performance comparisons of AODV, secure AODV and adaptive secure AODV routing protocols in free attack simulation environment. Eur J Sci Res 32(3):430–443. ISSN 1450–216X
11. Kannhavong B, Nakayama H, Nemoto Y, Kato N (2007) A survey of routing attacks in mobile ad hoc networks. IEEE Wirel Commun 14(5):85–91
12. Abusalah L, Khokhar A, Guizani M (2008) A survey of secure mobile ad hoc routing protocols. IEEE Commun Surv Tutor 10(4):78–93
13. Zhang C, Zhu X, Song Y, Fang Y (2010) A formal study of trust- based routing in wireless ad hoc networks. Proceedings IEEE, INFOCOM
14. Weerasinghe H, Fu H (2008) Preventing cooperative black hole attacks in mobile ad hoc networks: simulation implementation and evaluation. International journal of software engineering and its applications
15. Burmester M, de Medeiros B (2009) On the security of route discovery in MANETs. IEEE Trans Mobile Comput 8(9):1180–1188

16. Mahapatra RP, SM IACSIT, Katyal M (2010) Taxonomy of routing security for ad-hoc network. Int J Comput Theory Eng 2(2):303–307
17. Olsr, RFC. http://www.ietf.org/rfc/rfc3626.txt?number=3626
18. Perkins CE, Bhagwat P (1994) Highly dynamic destination sequenced distance vector routing (DSDV) for mobile computers. In: ACM SIGCOMM'94, 1994
19. C. E. Perkins and E. M. Royer, "Ad-Hoc On-Demand Distance Vector Routing", Proc. 2nd IEEE Wksp. Mobile Computer Systems and Applications, 1999, pp. 90–100
20. Hu Y, Perrig A, Johonson DB (2002) Ariadne: a secure on-demand routing protocol for ad hoc networks. In: Proceedings of ACM MobiCom '02, 23–26 Sept 2002

Analysis of Different Mobility Models for Ad Hoc On-Demand Distance Vector Routing Protocol and Dynamic Source Routing Protocol

Gaurika Talwar, Hemika Narang, Kavita Pandey and Pakhi Singhal

Abstract A mobile ad hoc network (MANET) is a network consisting of a set of mobile nodes that are connected by wireless links, communicating with each other without any centralized control or established infrastructure. Since mobility patterns may play a significant role in determining the protocol performance, it is desirable for mobility models to emulate the movement pattern of targeted real life applications in a reasonable way. Otherwise, the observations made and the conclusions drawn from the simulation studies may be misleading. Thus, it is necessary to study the performance of these mobility models. These models represent the movement of the mobile nodes, changes in their location, velocity and acceleration over time. In this paper, we study different mobility models and their performance over Ad Hoc On-Demand Distance Vector Routing Protocol(AODV) and Dynamic Source Routing(DSR) routing protocols. There is an attempt to provide an overview of the performance of various mobility models based on different parameters including throughput, packet delivery ratio, end-to-end delay, and jitter and routing load.

G. Talwar · H. Narang (✉) · K. Pandey · P. Singhal
Department of Computer Science, Jaypee Institute of Information Technology, Noida, India
e-mail: narang.hemika@gmail.com

G. Talwar
e-mail: gaurika_talwar@hotmail.com

K. Pandey
e-mail: kavita.pandey@jiit.ac.in

P. Singhal
e-mail: singhalpakhi@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_57,

1 Introduction

Mobile ad-hoc networks (MANET) or “short live” networks operate in the absence of fixed infrastructure. They offer quick and easy network deployment in situations where it is not possible otherwise. MANET is an autonomous system of mobile nodes (MNs) connected by wireless links; each node operates as an end system and a router for all other nodes in the network. Nodes in MANETs are free to move around; therefore the topology is dynamic in nature. The routing protocol controls how nodes decide the way to route packets between the MNs in MANET. There are two types of routing protocols, proactive and reactive routing protocols. The proactive routing protocol maintains fresh lists of destinations and their routes by periodically distributing routing tables throughout the network whereas the reactive routing protocol finds a route on demand by flooding the network with Route Request packets. Mobility models represent the movement of MNs, their changing location, velocity and acceleration. The performance of any protocol depends on the duration of interconnections between two nodes transferring data and nodes of a data path containing n-nodes. The mobility of MNs affects the number of average connected paths, further affecting the performance of the routing protocol. For experimental studies, two reactive routing protocols Ad Hoc On-Demand Distance Vector Routing Protocol (AODV) and Dynamic Source Routing (DSR) are considered. The performance of each mobility model is evaluated on the basis of end to end delay, routing load, packet delivery ratio, throughput and jitter.

The above discussion leads us to evaluate the performance of eleven different mobility models with the help of the two routing protocols. We have increased the number of mobility models, which have comparable results and also we can conclude about the best model amongst eleven models.

Section 2 discusses the work already done by various researchers related to Mobility models and routing protocols. Section 3 provides an overview of eleven mobility models. Section 4 gives a brief description about the two different routing protocols being used for analysis. Section 5 presents simulation and the result details. Conclusion along with future work is given in Sect. 6.

2 Related Work

An overview of the current research status of mobility models and their analysis was provided by Fan et al. [2]. Mobility models with unique characteristics such as temporal dependency, spatial dependency or geographic restriction was discussed by Fan et al. [2].

The effects of mobility models on the performance of two routing protocols DSR and Destination-Sequenced Distance-Vector (DSDV) [6] was studied by Bhavyesh et al. [3]. Random Waypoint, Group Mobility, Freeway and Manhattan models were selected for comparison. Comparison was done by varying node densities and number of hops. The results showed that performance of the routing protocols varied for mobility models with respect to node densities and length of data paths. Fahim et al. [8] did the comparison of routing protocols like AODV, DSR, Dynamic MANET On-demand Routing Protocol (DYMO) [7] (Reactive protocols) and Optimized Link State Routing Protocol (OLSR) [12], DSDV (Proactive Protocols) against various mobility models such as Random WayPoint, Reference Point Group Model, Column Mobility Model. Comparison was drawn using Packet Delivery Ratio, Average Delay and Routing Load. The results showed the best protocol for a particular mobility model under a particular performance metric. The performance of DSDV, DSR, OLSR and AODV was evaluated by Nadir et al. [10] on Random Walk, Random Direction, and Random WayPoint models. Comparison was drawn using Packet Loss, Average end-to-end delay and Overhead by varying the mean speed. The conclusions provided the idea of choosing the routing protocol according to the mobility characteristics of underline scenario.

3 Mobility Models

For network simulation, researchers use two types of mobility models: traces and synthetic models. Traces are mobility patterns observed in real life systems. They provide accurate information for large number of nodes and long simulation period. Synthetic models can further be classified as follow:
Entity mobility models represent MNs with movements independent of each other.
Group mobility model represent MNs with movements dependent on each other.
This section discusses the eleven synthetic mobility models, which have been selected for evaluation, with detailed explanation for how they emulate real world scenario. The first seven models are entity models followed by four group models.

1. Random Walk Mobility Model

This model is based on random directions and speeds. MNs move from one location to another by randomly choosing a direction and speed from predefined ranges. If an MN reaches simulation boundary, it “bounces” to the simulation border.

2. Random Waypoint Mobility Model

This model includes pause time between change in speed and/or direction. At every instant an MN randomly chooses a destination and moves towards it with a speed chosen from a predefined range, [minspeed, maxspeed]. After reaching the destination, the node pauses for a defined time.

3. Random Direction Mobility Model

This model forces MNs to travel to the edge of the simulation area before changing direction and speed. The MNs choose a random direction in which it has to travel. The MNs then travel to the border of the simulation area in that direction.

4. Gauss-Markov Mobility Model

This model involves different levels of randomness via tuning parameters. Initially, each MN is assigned a speed and direction. At every instant, the speed and direction of each MN gets updated. The value of speed and direction at the nth instance is calculated based on the value of speed and direction at (n − 1)th instance.

5. Probabilistic Version of the Random Walk Mobility Model

This model utilizes a set of probabilities to determine the next position of an MN. Also, it utilizes a probability matrix that defines the probabilities of a node moving forwards, backwards, or remain at the same position in both x and y direction.

6. Static Model

In this model, nodes are static and homogeneously distributed over the simulation area. For non-homogeneous node distributions, attraction points are defined or the simulation area is divided into areas with different node densities along x-axis.

7. Manhattan Grid Model

This model uses a grid road topology. It produces the movement pattern of MN on streets defined by maps. Here, all MNs on same street have identical speed. MN is allowed to move along horizontal and vertical streets on the simulation area.

8. Column Mobility Model

In this model, the MNs form a line and move forward uniformly in a particular direction. The column of reference points picks a random direction and speed. The nodes follow their reference point across the map. Based on “maxDist” i.e. the maximum distance a node can move, their random movement is determined.

9. Nomadic Community Mobility Model

In this model, MNs move together from one location to another. Each group has an invisible reference node to be followed around the simulation. As the reference node changes, MNs travel to a new location and start roaming. Their roaming is defined by picking random locations within roaming radius of the reference point.

10. Pursue Mobility Model

This model represents MNs tracking a single targeted node. It uses Random Waypoint with no pause time to move towards the specified target. Each MN determines its new location based on their old location, specified target’s movement and a random vector, where random vector is a random offset for each node.

11. Reference Point Group Mobility Model

This model represents the random motion of a group of mobile nodes as well as the random motion of each individual mobile node within the group. It includes the possibility to have dynamic groups. As the individual reference point move, their locations are updated according to the group’s center. When a node comes into the area of another group, it changes its group with a specified probability.

In the next section, the routing protocols selected for evaluating these mobility models discussed above will be discussed.

4 Routing Protocols

In order to understand the impact of mobility models on the performance of routing protocols, reactive routing protocols: AODV and DSR have been used. They function on-demand rather than proactively thus creating the route as per the requirement by flooding the network with route request packets. This section gives an overview of these two routing protocols that have been used for evaluation.

1. Ad Hoc On-Demand Distance Vector Routing Protocol(AODV)

It is a very famous protocol and therefore is preferred by the researchers. A source that intends to reach the destination floods the network with a route request (RREQ) packet to search all routes to the destination. On receiving the RREQ, each node creates a reverse routing entry for the source. The node also checks if it has an existing entry for the destination. If it has, a route reply (RREP) packet is generated and unicast back to the source along the reverse route request route.
Else, it rebroadcasts the first received RREQ and suppresses the duplicated ones. When the destination receives the first RREQ or RREQ coming from a shorter route, it sends a RREP back to the source. The nodes along the new routes, creates forward routing entries for the destination on receiving the RREPs. As routes are formed on demand network latency is high. The main advantage of this protocol is that route is established on demand and sequence numbers are used to determine the latest route to the destination. More details can be seen in [1].

2. Dynamic Source Routing Protocol (DSR)

It is a routing protocol based on source routing where all the routing information is maintained and updated at MNs. It has two major phases, Route Discovery and Route Maintenance. It initiates Route Discovery to form a route when a transmitting computer requests one. To determine source routes, address of each device between the source and destination is collected. The learned paths are used to route packets containing the address of each device the packet will traverse. This may result in high overhead for long paths or large addresses. RREP is generated only if the message has reached the intended destination node. To return the RREP, the destination node must have a route to the source node. If the route is in the destination node's route cache, the route would be used. Else, the node will reverse the route based on the route record in the RREQ's header. In case of fatal transmission, the Route Maintenance Phase is initiated wherein the Route Error packets are generated at a node. The Route Discovery Phase is initiated again. More details can be seen in [5].

5 Simulations And Results

To analyze the impact of mobility models on AODV and DSR, all the simulations have been performed on NS2 2.35. We evaluated eleven mobility models against two reactive routing protocols. For each mobility model, scenarios were generated using Bonn Motion [4] and these scenarios were then exported to NS2 [13]. The simulation area used is 690 × 690 m^2. The number of nodes was varied from 10 to 200. All scenarios were simulated for 250 s. The different parameters used for comparison are discussed with the results as follows.

1. Packet Delivery Ratio(PDR)

It is the ratio between the number of data packets sent and the number of data packets received successfully. Figure 1 shows the result of PDR for different mobility models under the two routing protocol by varying the number of mobile nodes.

2. End to End Delay

It is the time taken for a packet to be transmitted across a network from source to destination. Figure 2 shows the result of the end to end delay for different mobility models under the two routing protocol by varying the number of mobile nodes.

3. Throughput

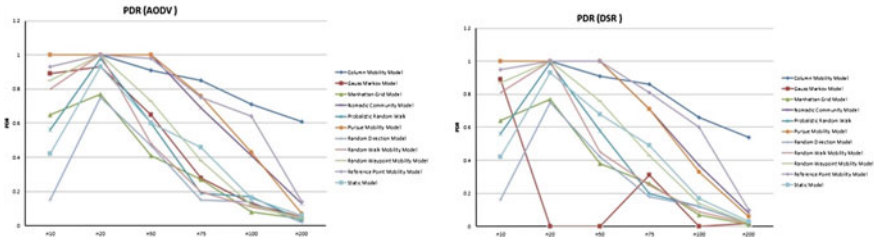


Fig. 1 PDR using AODV and DSR protocol

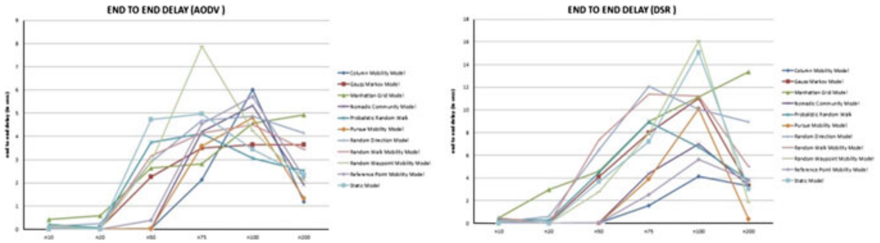


Fig. 2 End to end delay using AODV and DSR protocol

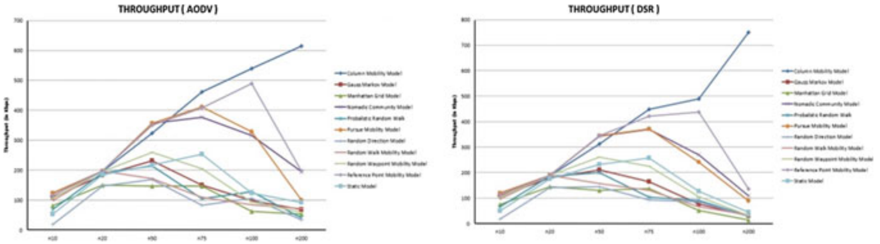


Fig. 3 Throughput using AODV and DSR protocol

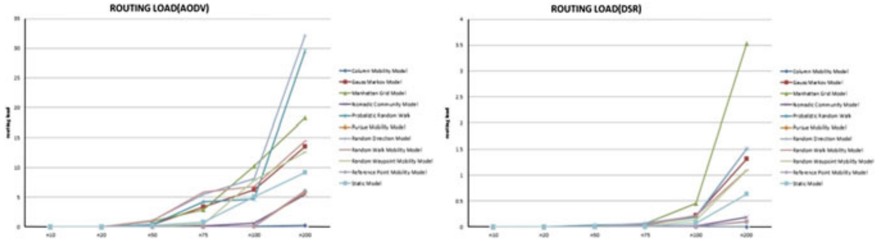


Fig. 4 Routing load using AODV and DSR protocol

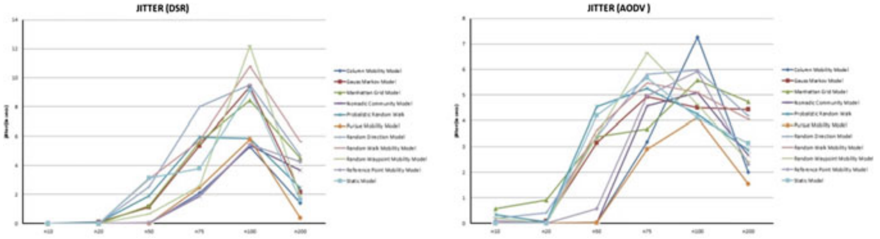


Fig. 5 Jitter using AODV and DSR routing protocol

Table 1 Performance of mobility models for AODV and DSR routing protocol with different parameters

	AODV	DSR
Throughput	Column mobility model	Column mobility model
Jitter	Pursue mobility model	Column mobility model
End to end delay	Column mobility model	Column mobility model
Packet delivery ratio	Column mobility model	Column mobility model
Routing load	Column mobility model	Column mobility model

It is the data transferred from one location to another in a given amount of time. Figure 3 shows the result of calculating the throughput for different mobility models under the two routing protocol by varying the number of mobile nodes.

Table 2 Performance of mobility models for DSR and AODV routing protocol with different number of mobility nodes

	10	20	50	75	100	200
Packet delivery ratio	Group models	Group models random walk random waypoint	Pursue reference point models	Column model	Column model	Column model
End to end delay	Pursue model	Pursue model	Pursue model	Column model	Column model	Pursue model
Throughput	Pursue model	Random walk	Pursue model	Column model	Column model	Column model
Routing load	All models	Group models, Gauss-Markov	Pursue column models	Pursue model	Column model	Column model
Jitter	Random direction model	Manhattan grid model	Column model	Nomadic model	Column model	Pursue model

4. Routing Load

It is the number of routing packets transmitted per data packets delivered at the destination. Figure 4 shows the result of Routing Load for different mobility models under the two routing protocol by varying the number of mobile nodes.

5. Jitter

Lower the jitter; better the performance of the mobility model in any routing protocol. Figure 5 shows the result of jitter for different mobility models under the two routing protocol by varying the number of mobile nodes

In case of AODV and DSR protocol Group models fared better than the entity models with respect to various performance parameters. From Table 1, in case of AODV 'Pursue Mobility model 'shows maximum performance with respect to Jitter and for the rest 'Column Mobility Model' shows perfect performance. In case of DSR, with respect to all parameters 'Column Mobility Model' fares the best. From Table 2, it is seen that **'**Pursue Model' shows best performance for few nodes (10–20), 'Pursue and Column Mobility Model' show almost equal performance for medium number of nodes (50–75) and 'Column Mobility Model' fares well with high number of nodes (100–200).

6 Conclusion and Future Work

The results and analysis show that performance of routing protocols varies with different mobility models. Therefore, we have considered various parameters to judge the performance of eleven mobility models with AODV and DSR and vary the number of mobile nodes from 10 to 200.

In general, we can conclude 'Group Model's performance with different protocols is better than the Entity models.

As Future work it can be specified that the same study can be conducted by varying other parameters such as the speed of the nodes. The scenario's can be further made realistic by in depth study of the application.

References

1. AODV (2012) http://moment.cs.ucsb.edu/AODV/
2. Bai F, Helmy A Survey of mobility models in wireless Ad Hoc networks. University Of Southern California, USA
3. Bhavyesh D (2007) Impact of node mobility on MANET routing protocols models. Mumbai University, India
4. BonnMotion http://net.cs.uni-bonn.de/wg/cs/applications/bonnmotion
5. DSR http://en.wikipedia.org/wiki/Dynamic_Source_Routing
6. DSDV http://www6.ietf.org/proceedings/44/slides/manet-thesis-99mar.pdf
7. DYMO http://ianchak.com/dymo/draft-ietf-manet-dymo-12.txt

8. Fahim M (2011) MANET routing protocol vs mobility model: a performance evaluation, National University of Sciences and Technology (NUST). Islamabad, Pakistan
9. Generating Traffic Pattern http://www.ece.iupui.edu/tutorials/ns2/
10. Nadir S (2008) Performance evaluation of multiple routing protocols using multiple mobility models for mobile ad hoc networks. Beihang University of Aeronautics and Astronautics Beijing, China
11. Ns-2 http://www.isi.edu/nsnam/ns/tutorial/nsscript7.html
12. OLSR http://www.ietf.org/rfc/rfc3626.txt
13. Simulation using BonnMotion http://chandra-ns2.blogspot.in/2009/01/how-to-run-bonnmotion-for-ns-2.html

NIZKPDS to Achieve Non-repudiation

S. Samundeeswari and V. S. Shankar Sriram

Abstract Digital signatures are electronic signatures that involve mathematical techniques for ensuring the authenticity of a digital message. It is mainly used for achieving non repudiation. There are a variety of digital signature schemes including the ELGAMAL based digital signatures. This contribution presents a novel Non-Interactive Zero Knowledge Protocol based Digital Signature (NIZKPDS) scheme. The proposed scheme mitigates the limitations of the ELGAMAL Scheme. A simulation based on the proposed scheme is presented using a simple Java application.

Keywords Digital signature · ELGAMAL digital signature · Non-repudiation · Zero knowledge protocol

1 Introduction

Authentication ensures the authenticity of users who intend to use the system.The difference between the Signature and the Digital Signature (DS) is the previous one is handwritten signatures for paper document and the later one is for electronic document. The aim of signature scheme is to ensure the authenticity as well as it protect the signature from forgery by third party. The main goals of the DS are

- To provide non-repudiation (means that later any one of the sending or receiving part will not deny the transaction),
- To avoid unforgeability (means that no other unauthorized user will pretend like a sender or receiver to any other party)

S. Samundeeswari (✉) · V.S. Shankar Sriram
School of Computing, SASTRA University, Kumbakonam, India
e-mail: samu_sivi@yahoo.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_58,

DS plays a vital role in E-commerce, E-business and Share Marketing. In the fast development of computers and communications, data encryption techniques and digital signature schemes are significant techniques to protect the sensitive information from any modification and replay attacks by the third party.
Diffie and Hellman (DH) [3] introduced the notion of DS based on RSA trapdoor function, in which each document has a separate distinct signature which leads to deterministic signature. Later Digital signatures developed against chosen message attack by weaker assumption by Goldwasser et al. [11] with the help of usage of permutations by LeinHarn et al. [1] by using trapdoor permutations Bellare [8] et al. and using one-way permutation and one-way function by Naor et al. and Rompel et al. [6, 9]. In all the above methods the same document can have one or more valid signatures. The use of one-way hash function is somewhat remarkable since it is easy for the privileged user to verify the signature, but it is difficult to verify for the others. Even though the invariant property has a number of valid signatures for the same document the DS leads tounpredictable nature of signature by the attacker using the implementation of trapdoor. The combination of Non-Interactive Zero knowledge Proofs (NIZKP) and the one-way function leads to a strong digital signature which is secure against chosen plain text attack was introduced by Bellare and Rompel et al. [6, 15]. We identified the relationship between the invariant digital signature and NIZKP to produce secure and strong DS which avoids both the forgery and non-repudiation.

The rest of the paper is organized as follows. Section 2 consists of related work of DS, in Sect. 3 the analysis of the ELGAMAL DS scheme and its security flaws are given, in Sect. 4 explains the need for NIZKP and our proposed system, in Sect. 5 we give illustrations for both ELGAMAL DS and our proposed NIZKPDS system with experimentation results and comparisons Sect. 6 concludes this paper.

2 Related Work

In the beginning of DS scheme RSA by Rivest et al. [12] and DSA [7] systems are used for the generation of the signature with the use of sender's secret key and the verifiers can check the validity of the signature by using the public key of the sender. Hence to provide the confidentiality of the plain text (original message) the Authenticated Encryption (AE) methods are used Lee et al. [16]. Using the AE, only the designated verifier can verify the signature and recover the message. To make verification process tuff by the third part, hash function was introduced, but this hash function will not reduce the size of the message because the message must be recovered from the cipher text. The message is divided in to number of message blocks if it is large and each block is signed, then the signcryption was introduced to handle large messages by Zheng et al. [18]. The difficulty in the large message transmission is the attacker may re-order or delete some cipher text blocks by Hwang et al. [4]. To overcome this AE scheme message linkage was proposed by Tseng et al. [17].

Wu et al. gave a clear explanation that signcryption can be used long messages and the AE schemes can be used for short messages. When non-repudiation occurs, in all the above schemes both the sender and the receiver needs to reveal the secret key. In order to avoid revealing of the secret keys during the execution steps of the transaction Zero Knowledge Protocols (ZKPs) are used. The ZKP does not reveal the secret key of either the sender or the receiver during the entire execution of the protocol hence the third party cannot learn any useful information, proposed by the researchers Araki et al. and Chang et al. [2, 5]. The important role of ZKP is to verify the validity of the DS without using the secret keys of the communicating parties.

3 Analysis of ELGAMAL Digital Signature Scheme

The ELGAMAL signature [10] scheme contains three algorithms namely, key generation, signature generation and signature verification. The signature generation part uses a one-way hash function h.

1. Key Generation $EK_g(1^n)$: The security parameter is 1^n for all ELGAMAL family key generation and the random oracle $EK_g(1^n)$ is used to generate large prime p and a generator g of order $p-1$. These two numbers are publicly known. The private key is selected by the signer from $x \in Z_P^*$ and computes the public key $y = g^x \bmod p$
2. Signature Generation: $ES(x.m)$: The signature generation part uses a one-way hash function h and the message m. The signer selects a secret key k randomly from $k \in Z_P^*$ and the $\gcd(k, p-1) = 1$, then computes $r = g^k \bmod p$ and r can be computed off-line. Using the private value x, the scheme computes s using the following equation $h(m) = sk + xr \bmod (p-1)$ where h is the one-way hash function. Thus $s = k^{-1}(h(m) - xr) \bmod (p-1)$ and the signature is $sig = (r, s)$
3. Signature Verification algorithm: $EV(y, m, sig)$: The verifier accepts the signature only if $g^{h(m)} = y^r r^s \bmod p$ otherwise rejects. The verification process performs the following steps.
    - $y^r r^s \equiv g^{xr} g^{kk^{-1}(h(m)-xr)} \equiv g^{h(m)} \bmod p$
    - $g^{xr+ks} \equiv g^{h(m)} \bmod p$
    - $xr + ks \equiv h(m) \bmod p - 1$

Intended readers may refer [13] for further information on ELGAMAL Scheme.

3.1 Security Flaw in ELGAMAL Signature Scheme [3]

In this section we show how the existential forgery and the new signature can be determined from the old signature in ELGAMAL signature scheme is explained.

- If the hash function is not used in ELGAMAL DS, the existential forgery is possible. The attacker chooses two integers u, v with $\gcd(v, p-1) = 1$ and sets the values of r and s by calculating $r = g^u A^v \bmod p$ and $s = -rv^{-1} \bmod (p-1)$ then finds $x = su \bmod (p-1)$. With these values of r and s, finds the verification congruence $y^r r^s \equiv y^r g^{su} y^{sv} \equiv y^r g^{su} A^{-r} \equiv g^x \bmod p$. The same calculation can be applied for hash function and is working.
- It is possible to generate a new signature from the old one. The condition $1 \le r \le p-1$ is crucial, if it is not required then it is possible to generate a new signature from the old signature by the following steps. To sign the another document m_1 the attacker computes the following from the old values and finds a new signature (r_1, s_1).

4 The Proposed Methodology

4.1 *Objective of the Proposed Work*

The main objective of the proposed system is to avoid the existential forgery in and the generation of the new DS from the old signature ELGAMAL signature scheme. The proposed scheme also consists of three phases such as Key generation, signature generation and signature verification and modifies the verification part of the ELGAMAL DS scheme by not revealing the secret values of both the verifier and sender, instead the proposed scheme only verifies the validity of the DS by NIZKP. The proposed system overcomes the security flaws of ELGAMAL DS such that it avoids learning of the secret key during the transaction by any intruder and also avoids the existential forgery.

4.2 *The Proposed Protocol*

4.2.1 Key Generation Steps

1. The signer chooses a large random prime p and a primitive root $g \bmod p$
2. The signer chooses a private value a randomly from the set $\{1, 2, \ldots, p-2\}$ and computes $A = g^a \bmod p$
3. Now the signer's private key is a and the public key is (p, g, A)

4.2.2 Signature Generation Steps

1. The document $x \in \{0, 1\}^*$ is signed by the signer using the publicly known collision resistant function $h : \{0, 1\}^* \rightarrow \{1, 2, \ldots, p-2\}$

2. A random number is chosen from the set $k \in \{1, 2, \ldots, p-2\}$ by the signer which is prime to $p-1$
3. Signer computes $r = g^k \bmod p$, $s = k^{-1}(h(x) - ar) \bmod (p-1)$ where k^{-1} is the inverse of the k modulo $p-1$
4. The signature of the document x is the pair of (r, s)

4.2.3 Verification Steps

1. Prover computes $q = A^r r^s \bmod p \; and \; sends$ to the verifier.
2. Verifier chooses challenge $e \; from \; the \; set \{0, 1\}$ and send to the prover
3. Then prover sends the response based the e value, if the e value is 0 then prover calculates $n = A^r r^s$ and sends to the verifier.
4. Now verifier checks the value by calculating $n > p$ and $g^{h(m)} \equiv n \bmod p = q.$
5. If the e value is 1 then prover calculates $lando, l = A^r \bmod p$ and $o = r^s \bmod p$ and sends these two ($l \; and \; o$) values to the verifier.
6. Then the verifier checks $1 \leq l, o < p \; and \; g^{h(m)} \equiv l * o \bmod p = q.$

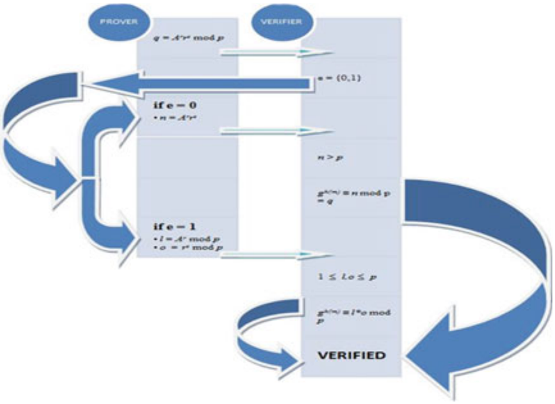
If these verifications are true then the verifier will be convinced by the prover.

The above Fig. 1 shows the verification process of the proposed system.

4.3 Security Analysis of NIZKPDS

During the signature verification, the verifier is not using the original signature (r, s) to verify the authenticity of the sender, instead the verifier checks the validity of the sender's signature by giving challenge $e(e \; from \; the \; set \; \{0, 1\})$ and getting response from the sender. If the challenge e is 0 ,the sender gives the response as

Fig. 1 Verification steps



$n = A^r r^s$ else if the challenge e is 1,sender gives the two response values l and o, $l = A^r \bmod p$ and $o = r^s \bmod p$. The verifier verifies the signature by the form of $AB(\bmod p) = (A(\bmod p) * B(\bmod p) \bmod p)$ of the sender. Hence the proposed NIZKPDS scheme avoids the existential forgery, and also avoids computing the new signature from the old signature values.

4.3.1 Proof of Completeness

At step 2, Verifier sends a random Value e and in turn prover computes depending on e, $n = A^r r^s$ or $l = A^r \bmod p$ and $o = r^s \bmod p$ and sends back to verifier. And verifier verifies $g^{h(m)} \equiv n \bmod p = q$ and $g^{h(m)} \equiv l * o \bmod p = q$. Thus verifier will verify the prover's signature (r,s) without knowing the values of r and s.

4.3.2 Proof of Soundness

As in step 1, prover generates $q = A^r r^s \bmod p$ and at step 2 prover generates $n = A^r r^s$. By assumption prover can not calculate n based on the value of q. Similarly in step 5 prover generates $l = A^r \bmod p$ and $o = r^s \bmod p$, provercan not generate feasibally any value of l and o without knowing (r,s), and satisfying the property $g^{h(m)} \equiv l * o \bmod p = q$.

5 Experimentation Results

5.1 Timing Analysis

The number of nodes taken for the ELGAMAL DS and the time required for the verification process is represented in the following Table 1.

The timing analysis of our proposed NIZKPDS system is compared with the ELGAMAL system using Java socket programming.The graphs are obtained as follows.

Table 1 Nodes taken for ELGAMAL DS

Number of nodes	Time in seconds
0.05	1
0.1	2
0.15	3
0.2	4
0.25	5
0.3	6

5.1.1 ELGAMAL System

The Fig. 2 shows the timing analysis of ELGAMAL DS based on the above Table 1.

5.2 *The Proposed NIZKPDS System*

The number of nodes taken for the proposed system and the time required for the verification process is represented in the following Table 2.
The timing analysis based on the Table 2 of the proposed system is implemented in the following Fig. 3.

5.3 *Comparison of ELGAMAL System and NIZKPDS System*

In the above Fig. 4, x-axis represents number of nodes authenticated and the y-axis represents the time taken for authenticating the respected number of nodes.

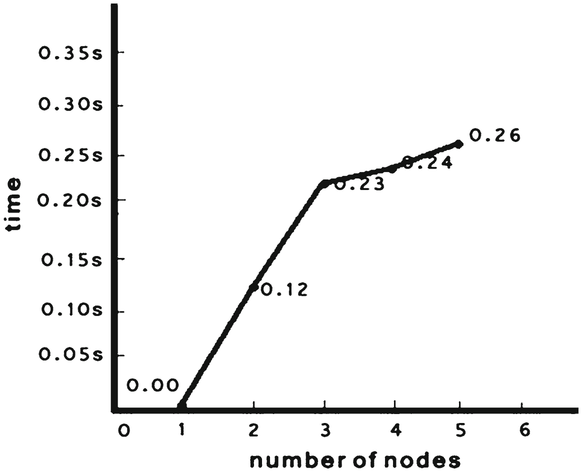


Fig. 2 Timing Analysis of ELGAMAL System

Table 2 Number of nodes taken for the proposed NIZKP DS

Number of nodes	Time in seconds
0.05	1
0.1	2
0.15	3
0.2	4
0.25	5
0.3	6

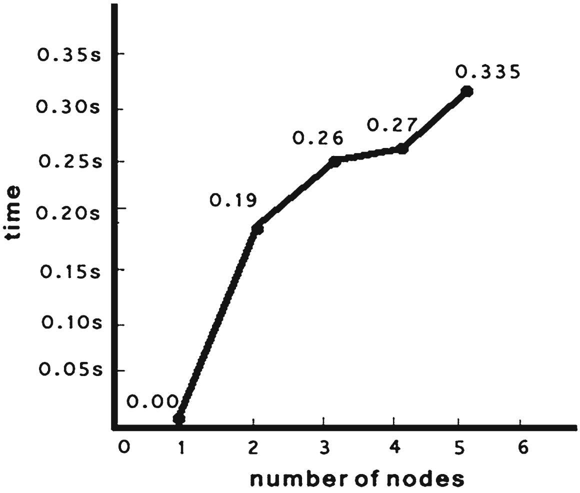


Fig. 3 Timing analysis of NIZKPDS system

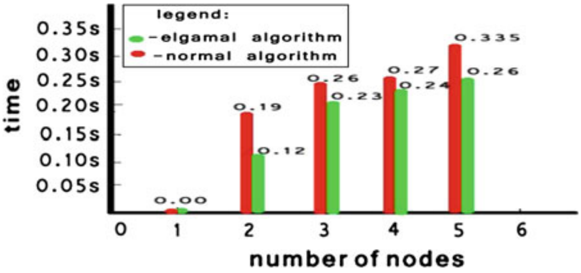


Fig. 4 Comparison of ELGAMAL and NIZKPDS

The time values obtained are the average of the values procured from multiple executions. From the experimental results it is observed that NIZKPDS performs better than the existing ELGAMAL DS.

The comparison of ELGAMAL system with the proposed NIZKPDS is done in terms of communicating with number of nodes asking for signing the document.

6 Conclusion

The Proposed NIZKPDS uses NIZKP protocol and hence it avoids the well known flaws of the ELGAMAL scheme like existential forgery and the new signature determination from the old signature.The experimental results proves that the NIZKPs can be used for the same.

References

1. Araki S, Uehara S, Imamura K (1999) The limited verifier signature and its application. IEICE Trans Fundam E82-A(1):63–68
2. Bellare M, Mical S (1988) How to sign given any trapdoor function, STOC 88. In: Proceedings of the 20th Annual Symposium on Theory of Computing, ACM, pp 214–233
3. Buchmann JA (2005) Introduction to cryptography, 2nd edn. Springer, New York
4. Bellare M, Goldwasser S (1989) New paradigms for digital signatures and message authentication based on non-interactive zero knowledge proofs. In: Proceedings of Crypto 89, pp 194–211
5. Chang Y-S, Wu T-C, Huang S-C (2000) ElGamal-like digital signature and multisignature schemes using self-certified public keys. Int J Syst Softw, Hindawi publications, 50:99–105
6. Diffie W, Hellman M (1976) New directions in cryptography. IEEE Trans Inf Theor IT-22: 644–654
7. ElGamal T (1985) A public-key cryptosystem and a signature scheme based on discrete Logarithms. IEEE Trans Inf Theor IT-3 1 (4):745–756
8. Goldwasser S, Micali S, Rivest R (1988) A digital signature scheme secure against adaptive chosen-message attacks. SIAM J Comput 17(2):281–308
9. Hwang SJ, Chang CC, Yang WP (1996) Authenticated encryption schemes with message linkages. Inf Process Lett 58(4):189–194
10. Lapidot D, Shamir A (1990) Publicly verifiable non-interactive zero-knowledge Proofs. In: Proceedings of Crpto 90, pp 356–367
11. LeinHarn A, JianRen B (2009) Design of DL-based certificateless digital signatures. J Syst Softw 82:789–793
12. Lee WB, Chang CC (1995) Authenticated encryption schemes without using a one way function. Electron Lett 31(19):1656–1657
13. Naor M, Yung M (1995) Universal one-way hash functions and their cryptographic applications, STOC 89
14. National Institute of Standards and Technology (NIST) (1992) The digital signature standard proposed by NIST. Commun ACM 35(7):36–40
15. Rompel J (1990) One-way functions are necessary and sufficient for secure signatures STOC 90
16. Rivest RL, Shamir A, Adleman L (1978) A method for obtaining digital signatures and public key cryptosystems. Commun ACM 21(2):337–348
17. Tseng YM, Jan JK, Chien HY (2003) Authenticated encryption schemes with message linkages for message flows. Comput Electr Eng 29(1):101–109
18. Zheng Y (1997) Digital signcryption or how to achieve cost (signature & encryption) $-$ cost (signature) $+$ cost

An Integrated Solution for Both Monitoring and Controlling for Automization Using Wireless Sensor Networks: A Case Study

M. Gnana Seelan and Ch. A. S. Murty

Abstract Temperature monitoring plays a major role in controlling it according to its varied conditions. This process is common in all critical areas like data centre, server rooms, grid rooms and other data communication equipped rooms. This is mandatory for each organization/industry to impart such process, as most of the critical data would be in Data centre along with their network infrastructure which having various electronic, electrical and mechanical devices are involved for data transmissions. These devices are very much depend on the environmental factors such as temperature, moisture, humidity etc., and also emit heat in the form of thermal energy when they are in functional. To overcome these heats, the Server/Data centre room(s) would be engaged with multiple (distributed) Air-conditioning (AC) systems to provide cooling environment and maintain the temperature level of the rooms. With distributed systems, it is difficult to control and maintain steady state in each corner of the room. In general, the manual control (through human involvement) for monitoring and controlling the temperatures of various distributed systems in such data centres. But it needs a continuous human interaction for monitoring the same which is required huge manpower to monitor the same on 60X60x24X7 (every second) basis. This may be time consuming and also prone to human error. The existing method of controlling and monitoring the Air-conditioning systems requires proper wiring technology and also frequent maintenance which could lead to high implementation and Maintenance cost. The proposed paper is the study of automization of monitoring and controlling temperature as per desired requirements with WSN network.

Keywords WSN · Distributed AC · Temperature control · Blue-tooth · IR circuit · Temperature monitoring · Data and grid room · HVAC

M. G. Seelan (✉) · C. A. S. Murty
Centre for Development and Advanced Computing(C-DAC),
Hyderabad, India
e-mail: mgseelan@cdac.in

C. A. S. Murty (✉)
e-mail: chasmurty@cdac.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_59,

1 Introduction

In the recent years, new technologies are evolved on automization of various processes. With that we understood that, introducing an automatic controls in terms of temperature monitoring and control using wired or wireless devices could be the best way to reduce the human interaction and human prone errors for controlling the systems. In addition there is a necessity to maintain control level for distributed systems.

Data centre or High performance computing labs such as Grid, Cluster or Server rooms are identified as critical Infrastructure of an organization. There is a need to maintain stable temperature for secure the data communication, secure from various physical damages with respect to various environmental threats. So, there is necessity to have a temperature control and monitoring system to maintain the stability of room temperature in the Data centre to increase productivity in-terms of life of various electrical and electronic components.

However, the monitoring and controlling temperature with in Data centre is very difficult as the flow of HVAC conditions are depended as per requirements of components in the Data centre and the requirements are different to different components and varies time to time [18]. Whenever the temperature exceeds the limit of threshold levels, many horrendous events can occur, such as server meltdown, loss of data and even fire may occur and may expect huge loss in terms of life, damages etc. These events also cost time and money to replace/repair damaged goods [17, 21–25, 27, 28].

1.1 Motivation

The monitoring and controlling various temperature levels for various components with in data centre would be difficult in manual monitoring and controlling as 24X7 basis and it would be better to have facility such as automization of such process on each minute and each second if possible. There are automatic controls for monitoring and controlling separately available for centralized systems and by keeping in view of distributed components, it would be required to have a combined automatic monitoring and controlling network for data centres. Temperature controls are commonly used in the world to maintain the temperature at certain level or ranges. There would be a seamless technology which can sense the temperatures on each second and compare with desired requirements and also inform to the automatic controls to control as per desired temperatures [2].

The main aim of the paper is to suggest and propose a low cost wireless sensor network with a low power IR remote sensor to control the distributed air conditioning system and integrated with automatic monitoring system. The main functions of the proposed system are: Continuous room temperature Monitor and Maintenance; to transmit signal to the server through receiver; Continuous controlling the distributed air conditioning system with respect to the temperature [19].

There are popular solutions for Automatic temperature monitoring and may be done through manually controlling the same by observing automatic temperature levels. But along with automatic monitoring, there is need for automatic controlling in distributed environment such as distributed Air Conditioning systems. This paper is discussing about the implementation of both automatic monitoring and controlling solutions as an integrated solution for monitoring and controlling various components through WSN networks [3–5, 7, 8, 14–16, 26].

2 Related Works

Globally, number of WSN solutions available for automization of temperature monitoring. But no solutions found for distributed systems. In view of above problem, we need a solution which could perform monitoring and instruct controller to control automatically as per required desired temperature levels. But we could not found a unique integrated solution which could perform both. From the literature survey for temperature control system and monitoring system some of the researchers and scientist mentioned as below.

The key motivation is from the Wireless Sensor Networks Research Group and is formed by researches and developers team throughout the world. Squid-Bee is an open-source WSN platform, where each Squid-Bee Mote takes *environmental parameters with its three sensors (humidity, temperature and lightness)* and sends through WSN [9].

Sridevi Veerasingam, Saurabh Karodi, Sapna Shukla and Mehar Chaitanya Yeleti on "Design of Wireless Sensor Network node on ZigBee for Temperature Monitoring" described the function of the wireless data logging system [20].

In continuation to the data logger the work by Ahmad Faris Bin Zulkifli on "Automatic Room Temperature Control With Security System" highlighted the importance of *automated temperature monitoring of a room and controlling with automated mechanism* [6].

The work by Vongsagon Boonsawat, Jurarat Ekchamanonta and Kulwadee Bumrungkhet on "XBee Wireless Sensor Networks for Temperature Monitoring" presented the function an embedded WS *capable of monitoring & managing N prototype system for temperature monitoring in a building* [1].

Over the past, many WSN systems have been developed and studied for numerous applications. On this regard lot of study been done to support the work and propose a integrated automatic temperature monitoring and controlling system.

3 Methodology

3.1 System Overview

The experimental integrated monitoring and controlling system is shown in Fig. 1. It consists of

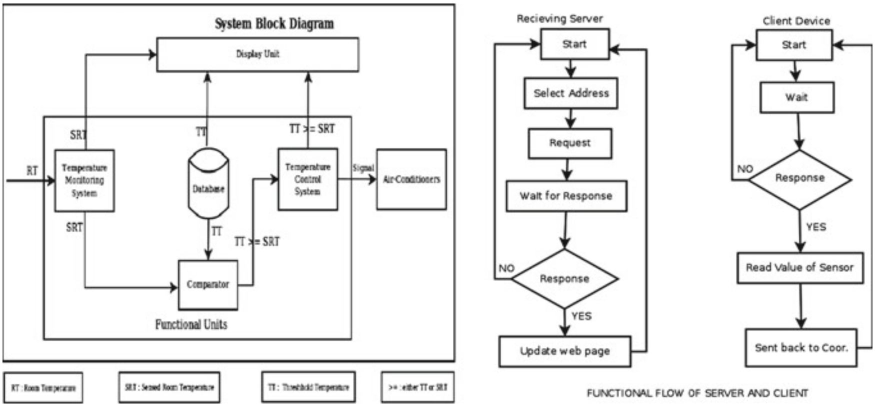


Fig. 1 Experimental setup & receiver and client functional diagram

(a) Temperature Monitoring System (TMS)
(b) Temperature Controlling System (TCS)
(c) Comparator and a Gateway

Temperature Monitoring System contains four parts as *(i) Client nodes which sense the surrounding temperature with respect to each nodes and wait the response from (ii) receiver which collects the sensing data from different sensors (client nodes) and forwards the same to (iii) the database server component which would be used stores the data and display using web user interface (iv) web server component for reports.*

Temperature Controller System contains three components as *(i) IR Remote connected to (ii) a computer which controls (iii) Air Conditioning(AC system) through Infrared communication.*

Comparator is having an web application to compare the threshold temperature(TT) with the sensing temperature(SRT). Serial Gateway of TMS is generally used for connecting receiver node with the database server.

3.2 Design and Development

The integrated temperature monitoring and controlling system has three major components as explained in the system overview. The Major components has been further divided as subsystem levels and been explained separately in the succeeding paragraphs.

Temperature Monitoring System: The experimental temperature monitoring system and its components are shown in Figs. 2 and 1.

Temperature Monitoring System contains four parts as:

(i) Client sensor nodes
(ii) Receiver node (Server node)

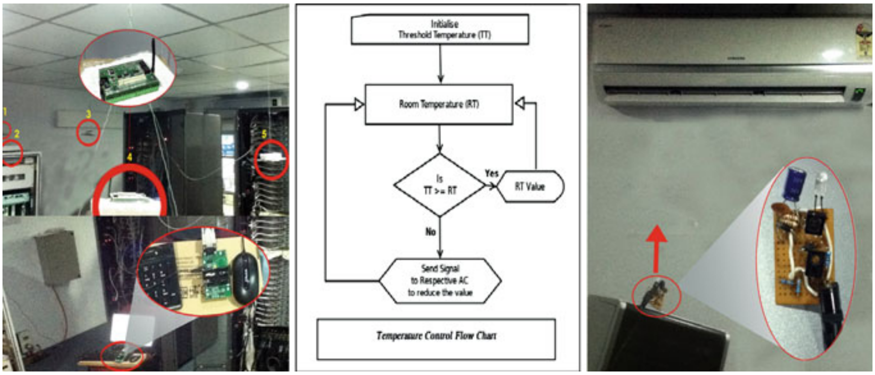


Fig. 2 Experimental temperature monitoring system, test condition & control system

(iii) The database server component
(iv) Web server component

Each of the subsystem explained briefly in the succeeding sections.

(i) Client Sensor nodes: The sensors *(nodes)* convert the analog temperature signal in terms of volts into degree Celsius according to the value. The client *(sensor)* node collecting the data through sensing technology and try to connect the server to upload as the same as immediately because of there is no storage space. Client sensor nodes which senses the surrounding temperature with respect to each nodes and wait the response from centrally placed receiver node. The function of the client nodes shown in the Fig. 1.

Wireless sensor network usually consists of a large number of nodes those are deployed in the sensing area and are equipped with different kinds of sensing, computation and communication units. These functional units enable WSN nodes to cooperatively collect, process, and transmit information to the receiver [19].

Client value with respect to the surrounding room temperature. The assumed values are to minimize the break down in the cluster room.

1. Node1 Value = 20.35 °C
2. Node2 Value = 22.15 °C
3. Node3 Value = 23.45 °C
4. Node4 Value = 26.35 °C
5. Node5 Value = 30.35 °C

(ii) Receiver node (server node): Receiver node *(Server node)* normally placed in the central location in a way to function using star topology. The function of the server node is to collects the sensing data from different sensors *(client nodes)* and forwards the same to the database server for updating the values with respect to the client nodes shown in Fig. 1. Receiver will respond to the all clients for updating the data.

(iii) Database server Component: Database server major function is to store each and every data received from the receiver. The database server also responsible for storing the thresh-hold Table 1 temperature of all the nodes. The data collected

Table 1 Thresh-hold temperature for room and setting for AC- No load on the cluster

Threshhold	Node-1	Node-2	Node-3	Node-4	Node-5	Values	AC-1	AC-2	AC-3	AC-4	AC-5
Minimum	17 °C	18 °C	22 °C	22 °C	25 °C	Maximum	24 °C	25 °C	23 °C	22 °C	21 °C
Maximum	23 °C	25 °C	27 °C	28 °C	32 °C	Minimum	23 °C	23 °C	22 °C	21 °C	20 °C

from the server node will be stored in the mysql/postgres/Sqlite database. We are using MySQL database for storing the data.

(iv) Web server component: In addition to the management of data there is a need to create reports of the collected data/output to analyse its performance, by web interface.To have web interface web server is required. We are using php *(lampp)* as a whole for hosting and retrieving data through web interface. The stored data would be displayed using php web user interface. The web servers other responsibilities are providing user friendly interface by which all the clients nodes temperature could be monitored, and could be utilized to create reports in the form of html etc.,.

Temperature Controlling System: Temperature controlling system has the major functionality of controlling AC with the usage of infrared remote. Infrared remote sends the infra red signal to the AC's as per controllers instruction. AC will set the value Table 1 with respect to the received infrared signal. In general, ACs are set with a range of values assigned to the voltage. The Fig. 2 demonstrates the practical implementation of temperature control system.

Temperature Controller System contains three components as:

(i) IR Remote circuit
(ii) Computer with control application
(iii) Air Conditioning(AC system)

(i) IR Remote circuit: Infrared mechanism will control as getting instruction from the controller. IR will send signal to the AC for respective values [10].

(ii) Computer with control application: To system is installed with web application using php (lampp) as a whole for hosting and retrieving data through web interface. The stored data would be displayed using php web user interface. The web servers other responsibilities are providing user friendly interface by which all the clients nodes temperature could be monitored, and could be utilized to create reports in the form of html etc.,.

(iii) Air Conditioning(AC system): The AC are deployed all over the room object is to control and stabilize the temperature inside the room.

Comparator and a Gateway: The comparator is an application configured in the monitoring system. The function of the application is to compare the sensed value collected by the receiver with the thresh-hold values of each nodes. There is a need to automate the comparison process by the comparator. The Serial gateway transmits the received data from the client nodes to the database server through serial interface. For experimental setup we used MySQL database to store the data's of each nodes and thresh-hold temperature.

Comparator is having an web application to compare the threshold temperature(TT) with the sensed room temperature(SRT). Serial Gateway of TMS is generally used for connecting receiver node.

4 System Specification

The system is considered as closed loop control system. Compared with the existing temperature monitoring and controlling system, this paper proposes an integrated solution based on wireless sensor network technology. This approach is to propose an automated & cost-effective realtime inhouse solution for high performance computing environment. The proposed system could communicate with individual nodes providing automated control to the distributed system.

Network Specification: The architecture is of star topology, receiver in the center to process collected datas of client nodes. It gathers sensor readings from all the client nodes.

Hardware Specification: A WSN node consists of some sensors as client and receiver, communication modules, Infrared Remote circuit, and a computer system for running application.

Sensors: Temperature sensors are having MICAz *(It is a 2.4 GHz Mote module used for enabling low-power, wireless sensor networks)*. The sensors support reliable wireless sensor network platform and ad-hoc mesh networking. They support tinyOS operating system and also provides reliability [11].

Other important components for the implementation of the systems are: Wireless Sensor Networking and Infra red with its functionalities; For the future work Know-how to develop applications using Blue-tooth [12] and Serial device, Blue-tooth and TCP Socket Programming on GNU/Linux and AVR Firmware Programming.

Software Specification:

Web-Server: We used html for web interface in the Centralized computer system. **Database**: We used MySQL server to run on the computer. The computer has LAMPP web server for Query MySQL command with PHP language accordingly.

System Limitation: As we are using multiple component for the experiment, there are various parameters need to be consider. The distance between the client nodes to the receiver are very narrow. Also there would be many line of site difference may occur between the nodes. The nodes may get over heated and send junk data to the receiver and which may collapse entire setup.

5 System Implementation

By considering 200 Square Feet area as pilot basis for experimental result, we implemented a prototype system of automated Temperature monitoring and control system based on Wireless Sensor Network (WSN). Although there are many wireless solutions as discussed in the Related Work section for monitoring or controlling, this system would be the solution for the distributed systems as shown in the figures above each system separately. It also satisfied the wireless communication standard specifications, and adapted to any data Centre and grid room with its extensive characteristics.

Considering the above, we deployed WSN system in a room which is having 40 node cluster and also not having proper air conditioning system. The system is designed in a way to monitor the stable temperature and control the air conditioning system for Clusters.

The Sensors *(Client nodes)* are placed in five different locations:

1. *Node1 placed 5 m from the back-end of the server rack*
2. *Node2 placed 4 m from the back-end of the server rack*
3. *Node3 placed 2 m besides from the server rack*
4. *Node4 placed 1 m from the front-side of the server rack*
5. *Node5 placed half a meter from the beck-end of the server rack*

Monitoring the sensitive data for various temperature changes due to exhaust of heat from physical nodes of cluster, the temperature varies if the jobs increase in the cluster.

The temperature inside the cluster room is always high and unstable and the human management was failed to control the temperature manually. Hence we proposed an automatic system is required to control the temperature within the cluster room by utilizing the temperature sensors.

6 Testing

We setup prototype with WSN supported in a room which has 40 node cluster, storage and other related devices. The actual control process is shown in the Fig. 2 explains about the test conditions. Individual test has been carried out with each component before integrating as a whole. The test has been done with placing the sensors nodes in different places of the specific room as discussed in the above section. The room was deployed with 5 different A/Cs.

The desired temperature for the normal function of the cluster should be maintained between 10 °C (50 °F) to 28 °C (82 °F) as per the load. Mostly all components such as computing nodes, and other networking equipment are designed to operate within a moderate narrow temperature ranges. To ensure reliable operation and the long life for components, we should ensure that the temperature stays within the limited band.

In this reason, there is a need to introduce laws of thermodynamics, as we are proposing a solution for temperature control which is normally in the form of heat energy.

It is from first law of thermodynamics the energy produced by a machine is proportional to the heat dissipation by the machine.

"In a thermodynamic process, the increment in the internal energy of a system is equal to the difference between the increment of heat accumulated by the system and the increment of work done by it" [13].

Likewise water flowing down-hill, as heat energy will naturally and automatically move from a hot place to a cooler place.

There is a need to maintain a temperature level for any component to function properly. The set temperature is called as thresh-hold temperature(TT). Each client node have different values with respect to the distance from the cluster. The thresh-hold temperature range is from 18 °C to 28 °C as the value will increase if the clients are placed near to the exhaust out of heat from the clusters. The temperatures vary from front side to backside of the clusters in general.

The each mote senses the respective surrounding temperature, and sends to the server mote to store in the computer server.

The collected sensed room temperature(SRT) need to compared with respective threshold temperature (TT). If the room temperature is less than or equal to the threshold then no action will be taken, if the temperature is greater than the threshold then a signal through IR will be sent to the respective air conditioner to change accordingly.

6.1 Component Testing

Every sensors are tested individually with the data error. It has been observed that whenever the temperature of the data centre increases it also affects the sensors. As the sensors are set with a functional voltage range. The receiver node will overload if all clients request receiver to upload the collected data.

Experimental Temperature measurements: As from law of thermodynamics explained in the above section it is understood that, heat energy is directly proportional to temperature, and also the following points are noted for the consideration while testing:

1. The distance from the cluster—placement of sensor nodes(clients).
2. The direction of the cluster with respect to the clients.
3. Start and end time of the jobs in the cluster.
4. Number of nodes utilized for running the job.
5. Distance from the AC, client nodes as well as IR remote.
6. Line of site of the client nodes with the receiver node.

We performed an experimental filed test for the temperature monitoring in the cluster room between 08:30 and 20:00.

The parameters considered to explain the scenarios:

1. *Temperature as (T) & Room Temperature as (RT)*
2. *Heat dissipation (Heat Energy) as (HE) & Error free Temperature as (EFT)*
3. *Sensed Room Temperature as (RST) & Sensor Temperature as (ST)*

Hence EFT = SRT-ST

The experiment was carried out for automization and manual control of the temperature control. For this we have taken reading and analyzed the data from the database are considered for scenario-1 *(Cluster with no jobs)*, Scenario-2 *(after the jobs started in the cluster)* and scenario-3 *(Cluster with full load)*. At the end we have explained the reason for the difference in the values.

As part of experiment, the client nodes are set to update the database for every 2 min interval with the data. The updated data in the database was so huge, it is difficult to compile and compare. Hence we considered the data for every 30 min interval for this paper from the database.

Normally the temperature varies inside the cluster room because of the heat emission of each server in the cluster as discussed above from law of thermodynamic. We observed that front side of the server is cooler than backside of the server node.

Scenario-1: **Cluster with no load**: The diagram for the cluster no loads (no jobs) running on the cluster shown in the Fig. 3. We also set the AC values in Normal to stable the room temperature and the details is shown in the Table 1 for the reference.

As the nodes are up and there were no jobs running hence the room temperature is normal and observation.

1. *The readings were taken from 08:30 to 19:30 h, for all five nodes.*
2. *The values of each nodes varied with respect to time and no relation to each other.*
3. *The peak value is observed at 16:00 h from Node5 with* 25.40 °C
4. The following observations for each nod were record as below:

Node1: *The minimal & maximum value observed as* 18.15 °C & 19.50 °C.
Node2: *The minimal & maximum value observed as* 19.25 °C & 21.30 °C.
Node3: *The minimal & maximum value observed as* 19.05 °C & 21.45 °C.
Node4: *The minimal & maximum value observed as* 19.25 °C & 21.45 °C.
Node5: *The minimal & maximum value observed as* 20.40 °C & 25.40 °C.

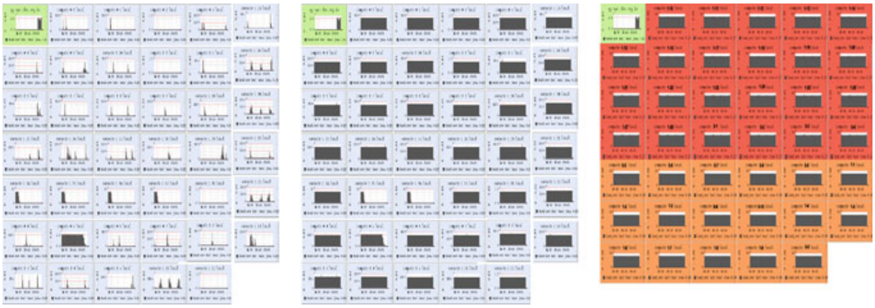


Fig. 3 Cluster with no load, initial load & full load

The above experimental output explains clearly that the temperature depends on the following:

1. *Continuous running of server increases the temperature and to manage the same requires manual adjustment of A/C's.*
2. *Temperature increases if no proper air conditioning facilities in the cluster room.*
3. *Temperature increases with the environmental heat such as external heats.*
4. *Temperature also increases according to the increase in the human intervention or movements in the cluster room.*
5. *The minimal value is from Node1 and the maximum value is from Node5.*

The results of variation in temperature levels with respect to the client with time are shown in Fig. 4.

Scenario-2: **Cluster with Start of load**: The Fig. 3 shows Initial stage of the cluster when jobs are started on the cluster.

The second scenario was tested with running number of jobs full load with running jobs in the cluster nodes. The reading has been considered after changing the AC values in each of the respective client nodes. If we adjust the ac values when temperature rises then it reduces and stabilizes. The experimental output shows that, when ever manual change is done it takes time as because of the following reasons:

Every time he/she has to set the value to the AC's. It may not worth full because the rise of temperature will not be slow. Human body is not addicted with high temperature situation, some times there could be sudden as well abnormal rise in the temperature. During the period human cannot control the same, which would lead to damage.

1. *The readings were available from 08:32 to 19:32 h, for all five nodes.*
2. *The peak value is observed at 17:02 h from Node5 with* 41.50 °C and
3. *The observation for each node separately noted after the start of the jobs in the cluster, the brief explanation for each node is given in the subsequent points:*

Node1: *The minimal & maximum value observed as* 21.10 °C & 24.25 °C.
Node2: *The minimal & maximum value observed as* 21.35 °C & 26.35 °C.
Node3: *The minimal & maximum value observed as* 22.15 °C & 29.55 °C.

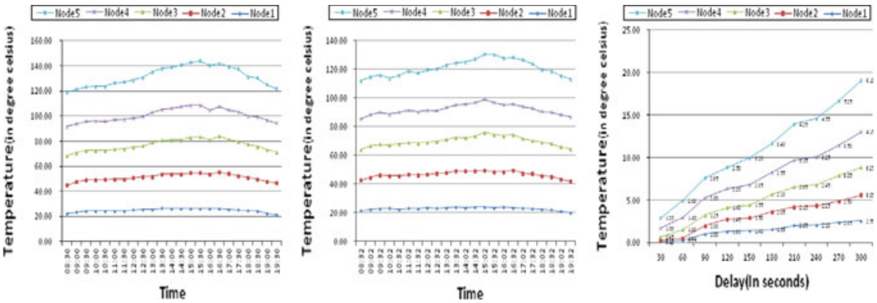


Fig. 4 The Initiat & after changing the value of AC in °C & temperature rise

Node4: *The minimal & maximum value observed as* 24.25 °C & 29.55 °C.
Node5: *The minimal & maximum value observed as* 29.10 °C & 41.50 °C.

As from above observation it is clear that the temperature increase with the load in the cluster and huge difference between minimum and maximum.

The results of variation in temperature levels with respect to the client with time are shown in Fig. 4.

Scenario-3: **Cluster with Full load**: The observation has been taken after the cluster with full load and is shown in the Fig. 3.

1. *The readings were available from 08:50 to 19:30 h, for all five nodes and was considered every 45 min/One hour with the condition.*
2. *The peak value is observed at 17:02 h from Node5 with* 41.50 °C and
3. *The observation for each node separately noted after the start of the jobs in the cluster, the brief explanation for each node is given in the subsequent points:*

Time delay for the control: Form above observation it was understood that controlling the AC by manual (*human*) control with respect to automatic control is having huge difference. We observed that, to control each node each second there is a need to monitor the deference in the temperature and other parameters. Human involvement will increase and slow the processes.

The formula for calculating the rise with respect to temperature is explained below.

From the Fig. 4 delay is directly proportional to temperature. Hence Rise in temperature = temperature + delayed temperature of each nodes

A huge difference between the minimum and maximum temperature value with respect to time was observed. Due to human intervention, there might be delays to control the temperature. For example, if delay is 30 s on Node1, the temperature is raining to 0.1 °C and if the same is about 300 s delay and the Temperature (T) raised to 2.55 °C. We considered the reading for every 30 s. Similarly, all the observations are shown in the Table 2.

With the above, it is analyzed that human or manual control is slower than the automization process. Any delay will lead to catastrophic damage.

Table 2 Comparison temperature in °C

Time/Nodes	30 s	300 s
Node-1	0.10 °C	2.55 °C
Node-2	0.25 °C	3.05 °C
Node-3	1.00 °C	3.25 °C
Node-4	1.30 °C	4.15 °C
Node-5	1.40 °C	6.10 °C

7 Summary of the Work

In this paper, we describe an integrated automatic temperature monitoring and controlling system using wireless sensor networks for distributed AC system. The study also describes about the existing centralized system for manual monitoring and controlling as separate or as a whole.

The concept describes about understanding the real time experiments in the data centre such as high performance computing environments. With our prototype we demonstrated the proposed system architecture can effectively satisfy the needs of any data centre or high performance computing environment with distributed AC systems.We also believe that wireless sensor networks could provide and optimal solution for the existing and futuristic conditions. The experimental study on the automization included the daily temperature levels, manual control and its impact with the control and automization.

8 Future Work

In our experiment, we used the common gateway which is a serial connector and limited to control single air conditioning sytem. In future, it could be utilized with various wireless technologies such as bluetooth for full pledged solution for different distributed or high performance environment like Grid computing, electrical control rooms/server room etc.

Acknowledgments We are grateful to the Members of Centre for Development of Advanced Computing (C-DAC), Hyderabad for providing their technical encouragement and supporting our efforts in the case study for providing a optimized solution for integrated automated temperature monitoring and controlling system.

References

1. Boonsawat V, Ekchamanonta J, Bumrungkhet K, Kittipiyakul S XBee wireless sensor networks for temperature monitoring. In: School of information computer and communication technology. Thani, Thailand
2. Choi W, Yoo W, Won S (2006) Development of automatic temperature control system in blast furnace. SICE- ICASE, Int Joint Conf 9(6):899–903
3. Dong X, Yang Y (2010) Study on household metering and temperature control in central air-Conditioning. In: IEEE ICIEA 2010. Zhongyuan University of Technology, China, pp 931–935
4. Lu H, Hong Z (2010) Design of Building Monitoring systems based on Wireless Sensors Network from college of Information Engineering, Wireless Sensor Network, vol 2. China, pp 703–709

5. Mcgranaghan M, Goodman F (2005) Technical and system requirements for advanced distribution automation. In: Proceedings of the 18th International Conference on Electricity Distribution. EPRI Report 1010915, Turin, Italy, pp 5–10. http://www.epri.com/abstracts/Pages/ProductAbstract.aspx?ProductId=000000000001010915
6. Faris A, Zulkifli B (2009) Automatic room temperature control with security system http://umpir.ump.edu.my/397/1/3315.pdf
7. Berger L (2003) Generalized multi-protocol label switching (GMPLS) signaling functional description. RFC 3473, Jan 2003
8. Gyongy IJ, Clarke DW (2006) On the automatic tuning ang adaptation of PID controllers. Control Eng Pract 14:149–163
9. http://www.sensor-networks.org/index.php?language=english\&page=the_group
10. http://www.extremecircuits.net/2010/04/ir-infrared-detector-circuit-diagram.html
11. http://www.openautomation.net/page/productos/id/22/title/MICAz-2.4-GHz
12. http://aopen.jp/tech/techinside/bluetooth.html
13. http://en.wikipedia.org/wiki/First_law_of_thermodynamics
14. http://public.web.cern.ch/Public/Welcome.html
15. http://ceaccp.oxfordjournals.org/content/8/3/104.full.pdf
16. http://www.freepatentsonline.com/6956516.html
17. http://www.hotfrog.com.au/Companies/Abacus-Instruments/Wireless-Temperature-Monitoring-System-181514
18. Jiang Y, Qin X (2000) Cooling metering and charging device in the fan coil unit air conditioning system. J Heating Vent Air Conditioning 30(6):48–55
19. Srikanth SV, Pramod PJ, Dileep KP, Tapas S, Patil MU, Sarat Chandra Babu N (2009) Design and implementation of a prototype Smart Parking (SPARK) system using wireless sensor networks. Centre for development of advanced, computing (C-DAC). WAINA09, Bradford, IEEEx-plore, pp 401–406
20. Veerasingam S, Karodi S, Shukla S, Yeleti MC (2009) Design of wireless sensor network node on ZigBee for temperature monitoring. In: ACT 2009, Department of instrumentation and control engineering. National Institute of Technology, Tiruchirappalli, India, pp 20–23
21. Veeraraghavan M, Zheng X, Huang Z (2006) On the use of connection-oriented networks to support grid computing. IEEE Commun Mag 44(3):118123
22. Wei D (2003) Discussion on the metering system of the centralized air conditioning system. Coal, Engineering, pp 57–58
23. Wu Q, Wang C (2009) Based on ZigBee technology for remote water meter reading system. Microprocessors 6(3):106–107
24. Xiaofang L, Yuntao Y (2009) A high accuracy temperature control system based on ARM9. In: ICEMI 2009, Qingdao University of Science and Technology, Shandong Province, China, pp 1795–1799
25. Yang H (2001) Cooling energy metering in the air conditioning system in Shanghai Xianxia Tennis Centre. Heating Vent Air Conditioning 31(5):39–42
26. Yao Y, Hu Y, Xu X (2001) Preliminary research on calculating method of cooling load in air-conditioned residence. J Build Energy Environ 8(50):1–11
27. Yuan L, Xiaohui Z (2009) Xiong construction, Wi-Fi-based wireless sensor network design and research. Modern Electron Technol 18:192–197
28. Zhang T, Lu K, Jue JP (2009) Differentiated contention resolution for QoS in photonic packet-switched networks

A Secure Image Steganography Technique to Hide Multiple Secret Images

S. Hemalatha, U. Dinesh Acharya, A. Renuka and Priya R. Kamath

Abstract Steganography is the science of "invisible" communication. The purpose of Steganography is to maintain secret communication between two parties. The secret information can be concealed in content such as image, audio, or video. This paper provides a novel image steganography technique to hide multiple secret images and keys in color cover image using Discrete Wavelet Transform (DWT). There is no visual difference between the stego image and the cover image. The extracted secret images are also similar to the original secret images. Very good PSNR (Peak Signal to Noise Ratio) values are obtained for both stego and extracted secret images. The results are compared with the results of other techniques, where single image is hidden and it is found that the proposed technique is simple and gives better PSNR values than others.

Keywords Steganography · DWT · MSE · PSNR · RGB · Luminance · Chrominance

1 Introduction

Information security is essential for confidential data transfer. Steganography is one of the ways used for secure transmission of confidential information. It contains two main branches: digital watermarking and steganography. The former is mainly used

S. Hemalatha (✉) · U. Dinesh Acharya · A. Renuka · P.R. Kamath
Manipal Institute of Technology, Manipal University, Manipal, Karnataka, India
e-mail: hema.shama@manipal.edu

U. Dinesh Acharya
e-mail: dinesh.acharya@manipal.edu

A. Renuka
e-mail: renuka.prabhu@manipal.edu

P.R. Kamath
e-mail: priyarkamath@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_60,

for copyright protection of electronic products while, the latter is a way of covert communication. Avoiding communication through well-known channels greatly reduces the risk of information being leaked in transit. Hiding information in a photograph of the company picnic is less suspicious than communicating an encrypted file.

The main purpose of steganography is to convey the information secretly by concealing the very existence of information in some other medium such as image, audio or video. The content used to embed information is called as cover object. The cover along with the hidden information is called as stego-object [5]. In this paper color image is taken as cover and two grey scale images are considered as secret information. Secret images and stego keys are embedded in the cover image to get stego image. The major objective of steganography is to prevent some unintended observer from stealing or destroying the confidential information. There are some factors to be considered when designing a steganography system [5]:

- Invisibility: Invisibility is the ability to be unnoticed by the human.
- Security: Even if an attacker realizes the existence of the information in the stego object it should be impossible for the attacker to detect the information. The closer the stego image to the cover image, the higher the security. It is measured in terms of PSNR.

$$\mathrm{PSNR} = 10\,\log\frac{L^2}{\sqrt{MSE}}\mathrm{dB} \tag{1}$$

where L = maximum value, MSE = Mean Square Error.

$$\mathrm{MSE} = \frac{1}{N}\sum\nolimits_{\mathrm{i}=1}^{N}|\mathrm{Xi}-\mathrm{Xi}'|^2 \tag{2}$$

where X = original value, X′ = stego value and N = number of samples.

High PSNR value indicates high security because it indicates minimum difference between the original and stego values. So no one can suspect the hidden information.

- Capacity: The amount of information that can be hidden relative to the size of the cover object without deteriorating the quality of the cover object.
- Robustness: It is the ability of the stego to withstand manipulations such as filtering, cropping, rotation, compression etc.

The design of a steganographic system can be categorized into spatial domain methods and transform domain methods [5]. In spatial domain methods, the processing is applied on the image pixel values directly. The advantage of these methods is simplicity. The disadvantage is low ability to bear signal processing operations. Least Significant Bit Insertion methods, Pallet based methods come under this category. In transform domain methods, the first step is to transform the cover image into different domain. Then the transformed coefficients are processed to hide the secret information. These changed coefficients are transformed back into spatial domain to get stego image. The advantage of transform domain methods is the high ability to face signal processing operations. However, methods of this type are computationally complex. Steganography methods using DCT (Discrete Cosine Transforms), DWT, DFT (Discrete Fourier Transforms) come under this category.

In this paper the secret images are embedded using DWT. The Wavelet Transform provides a time-frequency representation of the signal. DWT is used for digital images. Many DWTs are available. Depending on the application appropriate one should be used. The simplest one is Haar transform. To hide text message integer wavelet transform can be used. When DWT is applied to an image it is decomposed into 4 sub bands: LL, HL, LH and HH. LL part contains the most significant features. So if the information is hidden in LL part the stego image can withstand compression or other manipulations. But sometimes distortion may be produced in the stego image and then other sub bands can be used [5]. The decomposition of Lena image by 2 levels of 2D-DWT is shown in Fig. 1.

2 Related Work

Color images are represented in different color spaces such as RGB (Red Green Blue), HSV (Hue, Saturation, Value), YUV, YIQ, YCbCr (Luminance/Chrominance) etc. YCbCr is one of the best representations for steganography because the eye is sensitive to small changes in luminance but not in chrominance, so the chrominance part can be altered, without visually impairing the overall image quality much. Y is luminance component and Cb, Cr are the blue and red chrominance components respectively. The values in one color space can be easily converted into another color space using conversion formula [13]. Figure 2 shows the various components of Lena color image.

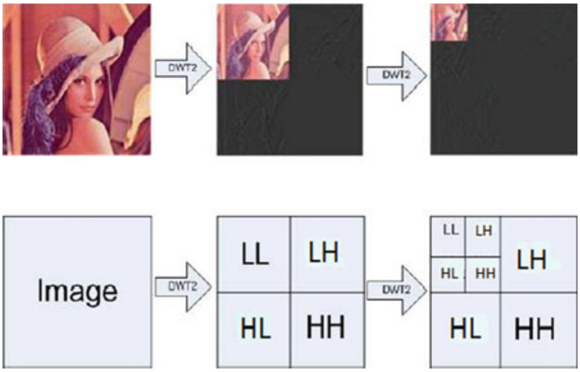


Fig. 1 2 Level 2D-DWT

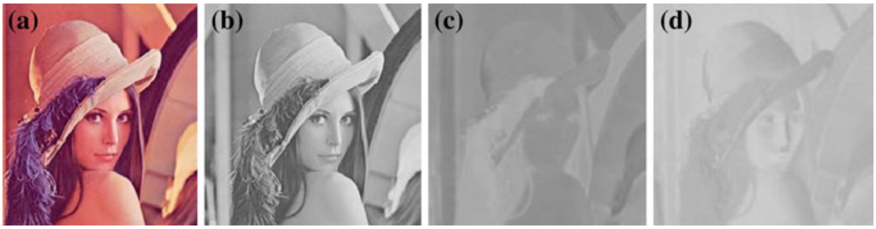


Fig. 2 **a** Lena, **b** luminance component Y of (**a**), **c** chrominance component Cb of (**a**), **d** chrominance component Cr of (**a**)

S. M. Masud Karim et al. [8] proposed a new approach based on LSB using secret key. The secret key encrypts the hidden information and then it is stored into different position of LSB of image. This provides very good security.

Xie Qing et al. [14] proposed a method in which the information is hidden in all RGB planes based on HVS (Human Visual System). This degrades the quality of the stego image.

In the method proposed by Sunny Sachdeva et al. [11] the Vector Quantization (VQ) table is used to hide the secret message which increases the capacity and also stego size.

The method proposed by Rong-Jian Chen et al. [1], presents the novel multi-bit bitwise adaptive embedding algorithm for data hiding by evaluating the most similar value to replace the original one.

Sankar Roy et al. [9] proposed an improved steganography approach for hiding text messages within lossless RGB images which will suffer from withstanding the signal processing operations.

Minimum deviation of fidelity based data embedding technique has been proposed by J. K. Mandal et al. [7] where two bits per byte have been replaced by choosing the position randomly between LSB and up to fourth bit towards MSB. A DWT based frequency domain steganographic technique, termed as WTSIC is also proposed by the same authors, [6] where secret message/image bits stream are embedded in horizontal, vertical and diagonal components.

Anjali Sejul et al. [13] proposed an algorithm in which binary images are considered to be secret images which are embedded inside the cover image by taking the HSV (Hue, Saturation, Value) values of the cover image into consideration. The secret image is inserted into the cover image by cropping the cover image according to the skin tone detection and then applying the DWT. In this method the capacity is too low.

Saeed Sarreshtedari et al. [12] proposed a method to achieve a higher quality of the stego image using BPCS (Bit Plane Complexity Segmentation) in the wavelet domain. The capacity of each DWT block is estimated using the BPCS.

Saddaf Rubab et al. [10] proposed a complex method using DWT and Blowfish encryption technique to hide text message in color image.

In the paper by Kapre Bhagyashri et al. [4] a new singular value decomposition (SVD) and DWT based water mark technique is proposed in full frequency band in YUV color space.
Nabin Ghoshal et al. uses a steganographic scheme for colour image authentication (SSCIA) [2] where the watermark image is embedded using DFT.
The proposed work is the extension of our previous work [3] to color images in which the secret image is transmitted without actually embedding in the cover image. Only the key is hidden in the cover image. The steps for embedding are as follows:

- Obtain single level 2D DWT of the cover-image C and secret-image S.
- The resulting transformed matrix consists of four sub-bands CLL, CHL, CLH and CHH and SLL, SHL, SLH and SHH obtained by transforming images C and S respectively.
- The sub-images CLL and SLL are subdivided into non-overlapping blocks BCk1 ($1 \leq \mathrm{k}1 < \mathrm{nc}$) and BSi ($1 \leq \mathrm{i} < \mathrm{ns}$) of size 2×2 where nc, ns are the total number of non-overlapping blocks obtained from sub-images CLL and SLL respectively.
- Every block BSi, is compared with block BCk1. The pair of blocks which have the least Root Mean Square Error is determined. A key is used to determine the address of the best matched block BCk1 for the block BSi.
- Then inverse 2D DWT is applied to get the cover image C.
- The Key is then stored using one of the spatial domain techniques in the cover image C. The simplest of the spatial domain techniques is LSB insertion algorithm.
- The resultant image is the stego-image.

The secret image can now be extracted from the stego image by following the steps mentioned below:

- From the stego-image G, obtain the secret key K1.
- Transform the stego-image into single level 2D DWT.
- This transformation results in four sub-bands GLL, GHL, GLH and GHH.
- Divide the sub-band image GLL into 2×2 non-overlapping blocks. The secret key K1 is used to obtain the blocks that have the nearest approximation to the original blocks in secret image.
- The obtained blocks are then rearranged to obtain the sub-band image SLLnew. Assuming SHLnew, SLHnew, SHHnew are zero matrices of dimension similar to SLLnew, the inverse 2D DWT is applied.
- The resultant image is the secret image that was originally intended to be embedded within the cover-image.

3 Proposed Method

In the proposed method, the cover is 256×256 color image. Two grey scale images of size 128×128 are used as secret images. In this approach, the following steps are performed for encoding:

- Represent the cover image C in YCbCr color space.
- Obtain single level 2D DWT of secret-images and Cb, Cr component of C.
- The resulting transformed matrix consists of four sub-bands corresponding to LL, LH, HL and HH sub bands.
- LL sub band of Cb is used to hide one secret image and LL sub band of Cr is used to hide another secret image using the same procedure used in our previous work [3] as described in Sect. 2.
- The two Keys corresponding to two secret images are then encrypted using simple exclusive or operation with a key and run length encoded and then hidden in the cover image using Least Significant Bit technique. Blowfish technique can be used for encryption. The resultant image is a stego-image G.

The secret images can now be extracted from the Cb and Cr components of the stego image using the same steps mentioned before for our previous work [3] in Sect. 2.

4 Experimental Results

The algorithm is tested in MATLAB. The results with different cover images and secret images are shown. Original cover and secret images are shown in Fig. 3. Two cover images "baboon" and "peppers" (Fig. 3a,b), each of size 256 × 256, are considered for testing the algorithm. The secret images considered are "earth", "football" and "moon" (Fig. 3c–e), each of size 128 × 128. The "football" and "earth" are embedded in "peppers". The resultant stego image is shown in Fig. 4a. The "earth" and "moon" are embedded in "baboon". The resultant stego image is shown in Fig. 4b. Extracted secret images from "peppers" are shown in Fig. 4c, d. Extracted secret images from "baboon" are shown in Fig. 4e, f. In all cases the average PSNR value of stego images is 44.7 dB. The PSNR values of the extracted secret images are also approximately 44.7 dB. The PSNR values in dB in all cases for stego and extracted secret images are tabulated in Tables 1 and 2 respectively. Table 3 compares the PSNR value of the stego image in the proposed method and that in the other four methods. In all these the cover image considered is "peppers" and the secret images used are of comparable sizes. The average PSNR value in the proposed method is much higher than that in the other methods.

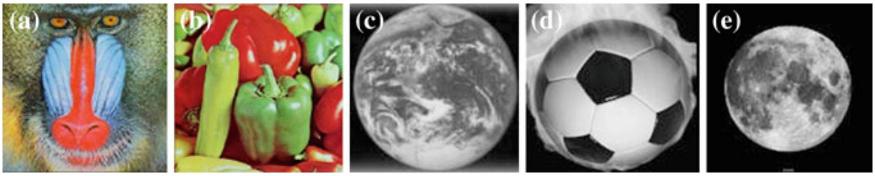


Fig. 3 Cover and secret images. **a** Cover (baboon). **b** Cover (peppers). **c** Secret (earth). **d** Secret (football). **e** Secret (moon)

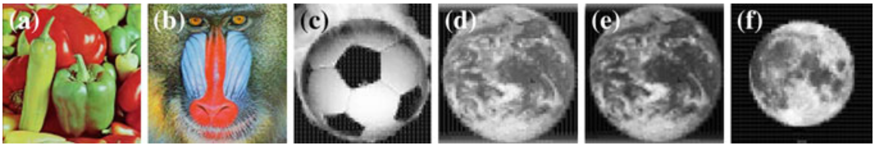


Fig. 4 Stego and extracted secret images. **a** Stego (football and earth as secret). **b** Stego (moon and earth as secret). **c**–**f** Extracted secret images: **c** football from peppers **d** earth from peppers, **e** earth from baboon, **f** moon from baboon

Table 1 PSNR (in dB) of the stego image

Cover image (256 × 256)	Secret images (128 × 128)	PSNR
Peppers	Football and earth	44.7
Baboon	Earth and moon	44.8

Table 2 PSNR (in dB) of the extracted secret image

Cover image (256 × 256)	Secret images (128 × 128)		
	Football	Earth	Moon
Peppers	44.6	44.7	
Baboon		44.8	44.8

Table 3 Comparison of PSNR (in dB) of the stego image in different methods

Technique	PSNR
STMDF [7]	39.6
WTSIC [6]	42.4
DWT-SVD [4]	36.6
SSCIA [2]	33.2
Proposed	44.7

5 Conclusion

In this paper, we observe that two secret images can be hidden in one color image and they can be regenerated without actually storing the image. This approach results in high quality of the stego-image having high PSNR values compared to other methods. However the disadvantage of the approach is that it is susceptible to noise if spatial domain techniques are used to hide the key. This can be improved if transform domain techniques are used to hide the key. The approach is very simple and the security level can be increased by using standard encryption techniques to encrypt the keys.

References

1. Chen RJ, Peng YC, Lin JJ, Lai JL, Horng SJ (2010) Novel multi-bit bitwise adaptive embedding algorithms with minimum error for data hiding. In: Proceedings of 4th international conference on network and system security (NSS 2010), IEEE Conference Publications, Melbourne, Australia, 1–3 Sept 2010, pp 306–311
2. Ghoshal N, Mandal JK (2011) A steganographic scheme for colour image authentication (SSCIA). In: Proceedings of international conference on recent trends in information technology (ICRTIT 2011), IEEE Conference Publications, Madras Institute of Technology, Chennai, India, 03–05 June 2011, pp 826–831
3. Kamath PR, Hemalatha S, Dinesh Acharya U, Renuka A (2012) High quality technique for image steganography using discrete wavelet transform. In: Proceedings of national conference on advances in computer science and information security (ACSIS'12), Manipal Institute of Technology, Manipal, India, 6–7 Jan 2012, pp 89–92
4. KapreBhagyashri S, Joshi MY (2011) All frequency band DWT-SVD robust watermarking technique for color images in YUV color space. In: Proceedings of IEEE international conference on computer science and automation engineering (CSAE), IEEE Conference Publications, 10–12 June 2011, pp 295–299
5. Katzenbeisser S, Petitcolas FAP (2000) Information hiding techniques for steganography and digital watermarking. Artech House, Inc., Boston
6. Mandal JK, Sengupta M (2010) Authentication/secret message transformation through wavelet transform based subband image coding (WTSIC). In: Proceedings of international symposium on electronic system design (ISED), IEEE Conference Publications, Bhubaneswar, India, 20–22 Dec 2010, pp 225–229
7. Mandal JK, Sengupta M (2011) Steganographic technique based on minimum deviation of fidelity (STMDF). In: Proceedings of second international conference on emerging applications of information technology (EAIT 2011), IEEE Conference Publications, 19–20 Feb 2011, pp 298–301
8. Masud Karim SM, Rahman MS, Hossain MI (2011) A new approach for LSB based image steganography using secret key. In: Proceedings of 14th international conference on computer and information technology (ICCIT 201I), IEEE Conference Publications, Dhaka, Bangladesh 22–24 Dec 2011, pp 286–291
9. Roy S, Parekh R (2011) A secure keyless image steganography approach for lossless RGB images. In: Proceedings of the international conference on communication, computing and security (ICCCS '11), ACM Publications, pp 573–576
10. Rubab S, Younus M (2012) Improved image steganography technique for colored images using wavelet transform. Int J Comput Appl 39(14):29–32
11. Sachdeva S, Kumar A (2012) Colour image steganography based on modified quantization table. In: Proceedings of 2012 second international conference on advanced computing and communication technologies (ACCT), IEEE Conference Publications, Rohtak, Haryana, India, 7–8 Jan 2012, pp 309–313
12. Sarreshtedari S, Ghaemmaghami S (2010) High capacity image steganography in wavelet domain. In: Proceedings of 2010 7th IEEE consumer communications and networking conference (CCNC), IEEE Conference Publications, Las Vegas, Nevada, USA, 9–12 Jan 2010, pp 1–5
13. Shejul AA, Kulkarni UL (2010) A DWT based approach for steganography using biometrics. In: Proceedings of the 2010 international conference on data storage and data engineering, IEEE Conference Publications, Bangalore, India, 9–10 Feb 2010, pp 39–43
14. Xie Q, Xie J, Xiao Y (2010) A high capacity information hiding algorithm in color image. In: Proceedings of 2010 2nd international conference on E-business and information system security (EBIISS2010), IEEE Conference Publications, Wuhan, China, 22–23 May 2010, pp 1–4

Source Code Analysis of a Connection-Oriented File Reader Server Socket Program in Java and Removal of the Security Vulnerabilities

N. Meghanathan

Abstract This paper presents the source code analysis of a file reader server socket program (connection-oriented sockets) developed in Java, to illustrate the identification, impact analysis and solutions to remove important software security vulnerabilities, which if left unattended could severely impact the server running the software and also the network hosting the server. The vulnerabilities studied are: (1) Resource Injection, (2) Path Manipulation, (3) System Information Leak, and (4) Denial of Service vulnerabilities. We analyze the reason for these vulnerabilities to occur in the program, discuss the impact of leaving them unattended, and propose solutions to remove each of these vulnerabilities from the program. We also analyze any potential performance tradeoffs (such as increase in code size and loss of features) that could arise while incorporating the proposed solutions on the server program. The proposed solutions are very generic in nature, and can be suitably modified to correct any such vulnerabilities in software developed in any other programming language.

Keywords Software vulnerabilities · Source code analysis · Resource injection · Path manipulation · System information leak · Denial of service · Unreleased resource · Network security

1 Introduction

In this paper, we focus on testing for software security using source code analysis (also invariably referred to as *static code analysis*). Static or source code analysis refers to examining a piece of code without actually executing it [3]. Source code analysis has been the principal means to evaluate the software with respect to functional, semantic

N. Meghanathan (✉)
Jackson State University, Jackson, MS, USA
e-mail: natarajan.meghanathan@jsums.edu

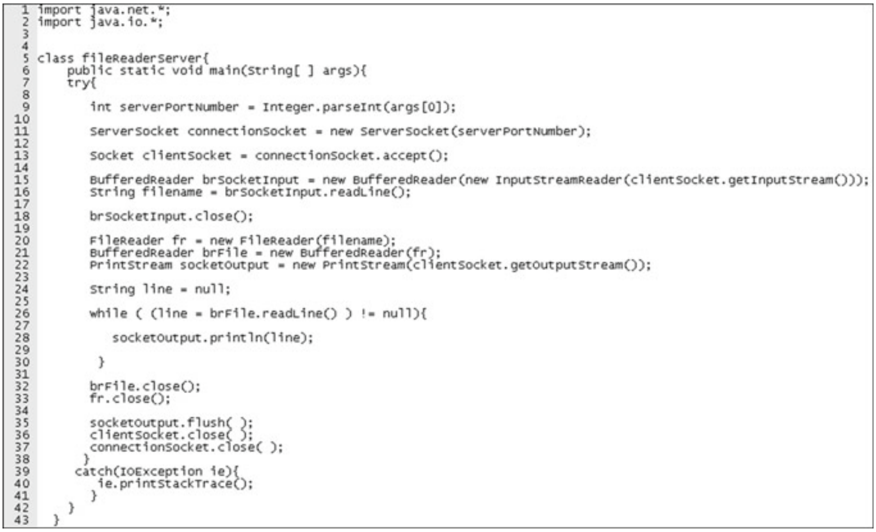
N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_61,

and structural issues including, but not limited to, type checking, style checking, program verification, property checking and bug finding [1]. On the top of these issues, the use of static code analysis to analyze the security aspects of software is gaining prominence. Static code analysis helps to identify the potential threats (vulnerabilities) associated with the software, analyze the complexity involved (in terms of increase in code size, development time, and code run time, etc) and the impact on user experiences in fixing these vulnerabilities through appropriate security controls [2]. Static code analysis also facilitates evaluating the risks involved in only mitigating or just leaving these vulnerabilities unattended—thus, leading to an attack, the consequences of such attacks and the cost of developing security controls and integrating them to the software after the attack has occurred [9]. In addition, static code analysis is also used to analyze the impact of the design and the use of the underlying platform and technologies on the security of the software [7]. For example, programs developed in C/Unix platforms may have buffer overflow vulnerabilities, which are very critical to be identified and mitigated; whereas, buffer overflow vulnerabilities are not an issue for software developed in Java. It would be very time consuming and often ineffective to manually conduct static code analysis on software and analyze the above issues as well as answer pertinent questions related to the security of software. One also needs to have a comprehensive knowledge of possible exploits and their solutions to manually conduct static code analysis.

Various automated tools have been recently developed to conduct static code analysis [6, 8]. In this paper, we illustrate the use of a very effective tool developed by Fortify Inc., called the Source Code Analyzer (SCA) [5]. The Fortify SCA can be used to conduct static code analysis on C/C + + or Java code and can be run in Windows, Linux or Mac platforms. The SCA can analyze individual program files or entire projects collectively. The analyzer uses criteria that are embedded into a generic rulepack (a set of rules) to analyze programs developed in a specific platform/ language. Users may use these generic rulepacks that come with the SCA or develop their own customized sets of rules.

2 Case Study on a Connection-Oriented File Reader Server Socket Program in Java

We present a case study on a file reader server socket program, based on connection-oriented sockets. We use the Fortify Source Code Analyzer to conduct the source code analysis of the file reader server program, implemented on a Windows XP virtual machine with the standard J2SE v.7 development kit. For simplicity, the server program is considered to serve only one client. The file reader server basically lets a client to read the contents of a file whose name or the path is sent by the client over a socket and the file is locally stored at the server. Figure 1 shows the original source code of the server program. We conduct source code analysis of



```
1  import java.net.*;
2  import java.io.*;
3
4
5  class fileReaderServer{
6      public static void main(String[ ] args){
7      try{
8
9          int serverPortNumber = Integer.parseInt(args[0]);
10
11         ServerSocket connectionSocket = new ServerSocket(serverPortNumber);
12
13         Socket clientSocket = connectionSocket.accept();
14
15         BufferedReader brSocketInput = new BufferedReader(new InputStreamReader(clientSocket.getInputStream()));
16         String filename = brSocketInput.readLine();
17
18         brSocketInput.close();
19
20         FileReader fr = new FileReader(filename);
21         BufferedReader brFile = new BufferedReader(fr);
22         PrintStream socketOutput = new PrintStream(clientSocket.getOutputStream());
23
24         String line = null;
25
26         while ( (line = brFile.readLine() ) != null){
27
28            socketOutput.println(line);
29
30          }
31
32         brFile.close();
33         fr.close();
34
35         socketOutput.flush( );
36         clientSocket.close( );
37         connectionSocket.close( );
38        }
39       catch(IOException ie){
40          ie.printStackTrace();
41         }
42     }
43  }
```

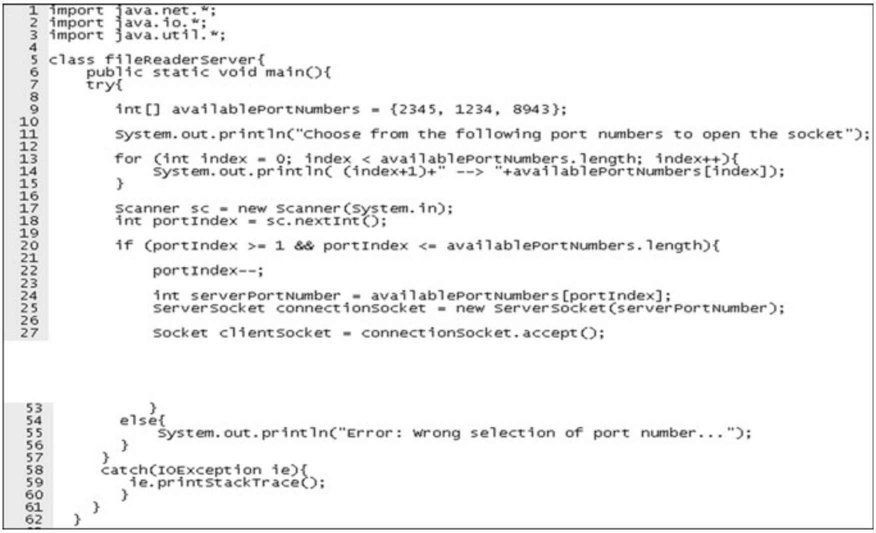
Fig. 1 Case study: original source code for the file reader server program

the file reader server socket program using the Fortify SCA and the output of all the issues identified.

2.1 *Resource Injection Vulnerability*

The Resource Injection vulnerability (a dataflow issue) arises because of the functionality to let the user (typically the administrator) starting the server program to open the server socket on any port number of his choice. The vulnerability allows user input to control resource identifiers enabling an attacker to access or modify otherwise protected system resources [3]. In the connection server socket program of Fig. 1, a Resource Injection vulnerability exist in line 11, wherein the program opens a socket on the port number whose value is directly input by the user. If the server program has privileges to open the socket at any specified port number and the attacker does not have such a privilege on his own, the Resource Injection vulnerability allows an attacker to gain capability to open a socket at the port number of his choice that would not otherwise be permitted. Thus, the program could even give the attacker the ability to transmit sensitive information to a third-party server.

In this section, we use the level of indirection approach to remove the Resource Injection vulnerability (refer to the modified code, especially lines 9 through 25 and 54–56, in Fig. 2). The user starting the server program is presented with a list of port numbers to choose from. Each valid port number is presented with a serial number and the user has to choose one among these serial numbers. If the user choice falls outside



```
import java.net.*;
import java.io.*;
import java.util.*;

class fileReaderServer{
     public static void main(){
     try{

         int[] availablePortNumbers = {2345, 1234, 8943};

         System.out.println("Choose from the following port numbers to open the socket");

         for (int index = 0; index < availablePortNumbers.length; index++){
              System.out.println( (index+1)+" --> "+availablePortNumbers[index]);
         }

         Scanner sc = new Scanner(System.in);
         int portIndex = sc.nextInt();

         if (portIndex >= 1 && portIndex <= availablePortNumbers.length){

              portIndex--;

              int serverPortNumber = availablePortNumbers[portIndex];
              ServerSocket connectionSocket = new ServerSocket(serverPortNumber);

              Socket clientSocket = connectionSocket.accept();




              }
          else{
               System.out.println("Error: Wrong selection of port number...");
          }
       }
       catch(IOException ie){
          ie.printStackTrace();
          }
     }
  }
```

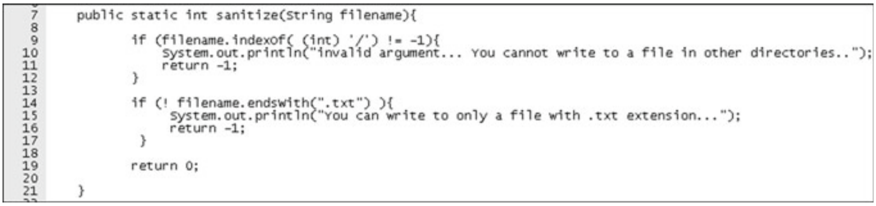
Fig. 2 Modification to the file reader server program to remove the resource injection vulnerability (fileReaderServer_1.java)

the valid range of these serial numbers, then the server program terminates printing a simple error message. The limitation is that the user no longer has the liberty to open the server socket at a port number of his choice. This is quite acceptable because often the server sockets are run on specific well-defined port numbers (like HTTP on 80, FTP on 21, etc) and not on arbitrary port numbers, even if the administrator wishes to run the server program on a port number of his choice.

2.2 Path Manipulation Vulnerability

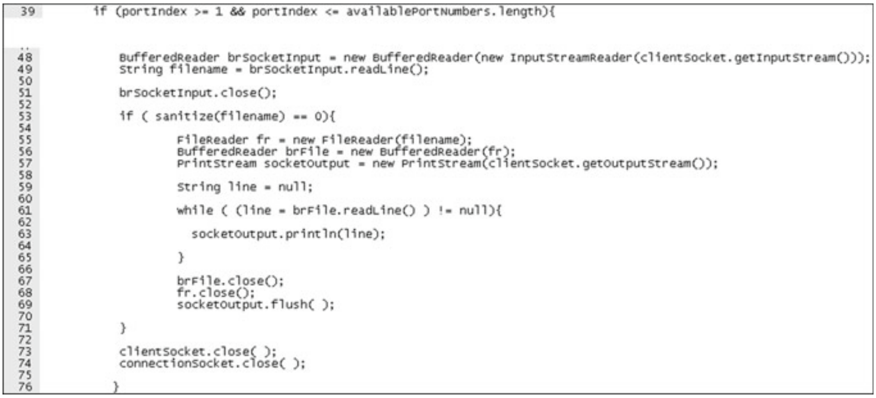
The Path Manipulation vulnerability occurs when user input is directly embedded to the program statements thereby allowing the user to directly control paths employed in file system operations [4]. In our file reader server program, the name or path for the file sent by the client through the socket is received as a String object at the server side, and directly passed onto the FileReader constructor (line 20 in Fig. 1). The practice of directly embedding a file name or a path for the file name in the program to access the system resources could be cleverly exploited by a malicious user who may pass an unexpected value for the argument and the consequences of executing the program, especially if it runs with elevated privileges, with that argument may turn out to be fatal. Thus, Path Manipulation vulnerability is a very serious issue and should be definitely not left unattended in a code. Such a vulnerability may enable an attacker to access or modify otherwise protected system resources.

We propose to use the approach of filtering user inputs using the blacklist/white list approach. It would not be rather advisable to present the list of file names to the client at the remote side—because this would reveal unnecessary system information to a remote user. It would be rather more prudent to let the client to send the name or the path for the file he wants to open, and we validate the input against a set of allowable and non-allowable characters. In this paper, we assume the file requested to be read is located in the same directory from which the server program is run, and that the file is a text file. Hence, the last four characters of the input received through the socket should be ".txt" and nothing else (thus, .txt at the end of the String input constitutes a white list). Also, since the user is not permitted to read a file that is in a directory other than the one in which the server program is running, the input should not have any '/' character (constituting a blacklist) to indicate a path for the file to be read. Here, we have implemented the solution of using white list and blacklist through the *sanitize*( ) method, the code for which is shown in Fig. 3. The modified file server program that calls the sanitize method to validate the filename before opening the file for read is shown in Fig. 4.



```java
public static int sanitize(String filename){

        if (filename.indexOf( (int) '/') != -1){
            System.out.println("invalid argument... You cannot write to a file in other directories..");
            return -1;
        }

        if (! filename.endsWith(".txt") ){
             System.out.println("You can write to only a file with .txt extension...");
             return -1;
         }

        return 0;
}
```

Fig. 3 Java code snippet for the sanitize method to validate the filename received through socket



```java
    if (portIndex >= 1 && portIndex <= availablePortNumbers.length){


        BufferedReader brSocketInput = new BufferedReader(new InputStreamReader(clientSocket.getInputStream()));
        String filename = brSocketInput.readLine();

        brSocketInput.close();

        if ( sanitize(filename) == 0){

                FileReader fr = new FileReader(filename);
                BufferedReader brFile = new BufferedReader(fr);
                PrintStream socketOutput = new PrintStream(clientSocket.getOutputStream());

                String line = null;

                while ( (line = brFile.readLine() ) != null){

                  socketOutput.println(line);

                }

                brFile.close();
                fr.close();
                socketOutput.flush( );

        }

        clientSocket.close( );
        connectionSocket.close( );

       }
```

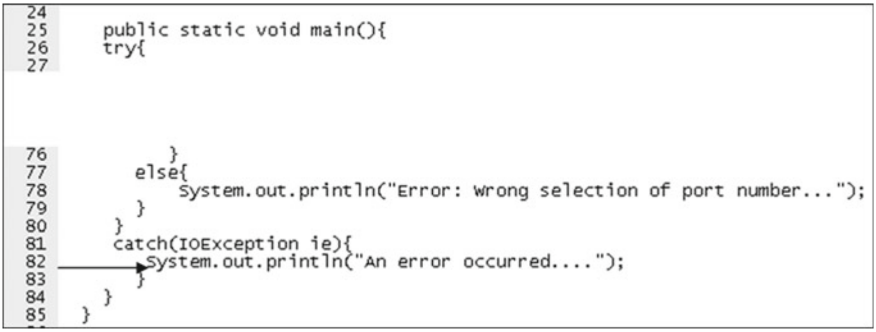
Fig. 4 Modified file reader server socket program to call the sanitize method to validate the filename before opening it to read (fileReaderServer_2.java)

2.3 System Information Leak Vulnerability

The "System Information Leak" vulnerability (a semantic issue) refers to revealing critical system data, program structure including call stack or debugging information that may help an adversary to learn about the software and the system, and form a plan of attack [10]. In our file reader server program (see Fig. 2), we observe that in line 82, the *printStackTrace*( ) method called on the object of the class *IOException* has the vulnerability to leak out sensitive system and program information including its structure and the call stack. While revealing the information about the call stack leading to an exception may be useful for programmers to debug the program and quickly as well as effectively trace out the cause of an error, the *printStackTrace*( ) method needs to be removed from the final program prior to deployment. A simple fix to this vulnerability is not to reveal much information about the error, and simply state that an error has occurred. The attacker, if he was contemplating to leverage the error information to plan for an attack, would not be able to gain much information from the error message. In this context, we remove the call to the *printStackTrace*( ) method from line 82 and replace it with a print statement just indicating that an error occurred. Figure 5 shows a modified version of the file reader server program.

2.4 Denial of Service Vulnerability

A 'Denial of Service' vulnerability (a semantic issue) is the one with which an attacker can cause the program to crash or make it unavailable to legitimate users [4]. Lines 49 and 61 of the file reader server socket program (as indicated in Fig. 4) contain the Denial of Service vulnerability, and this is attributed to the use of the readLine( ) method. It is always not a good idea to read a line of characters from a file through a program because the line could contain an arbitrary number of characters, without a



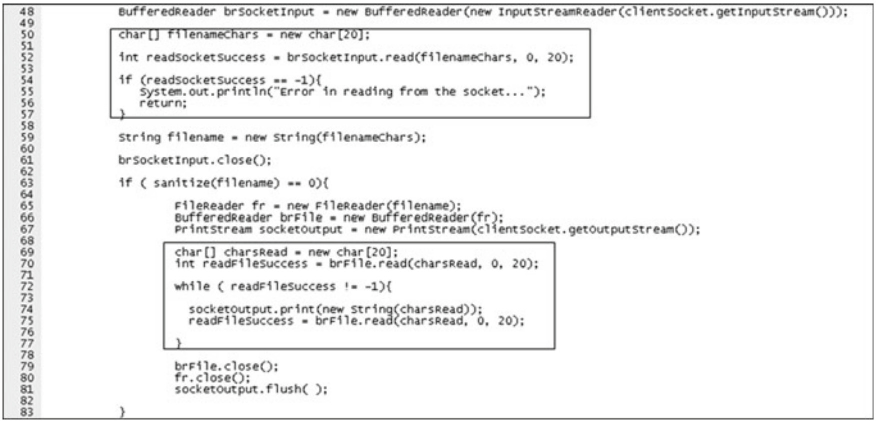
```
24
25      public static void main(){
26      try{
27

76              }
77          else{
78              System.out.println("Error: Wrong selection of port number...");
79          }
80        }
81        catch(IOException ie){
82            System.out.println("An error occurred....");
83          }
84      }
85  }
```

Fig. 5 Modified file reader server program to remove the system information leak vulnerability (fileReaderServer_3.java)

prescribed upper limit. An attacker could misuse this and force the program to read an unbounded amount of input as a line through the readLine( ) method. An attacker can take advantage of this code to cause an *OutOfMemoryException* or to consume a large amount of memory so that the program spends more time performing garbage collection or runs out of memory during some subsequent operation.

The solution we suggest is to impose an upper bound on the number of characters that can be read from the file and buffered at a time (i.e., in one single read operation). In this context, we suggest to use the *read*( ) method of the *BufferedReader* class that takes three arguments: a character array to which the characters read from the buffer are stored, the starting index in the character array to begin storing characters and the number of characters to be read from the buffer stream. In the context of lines 49 through 54 in the fileReaderServer_4.java program (boxed in Fig. 6), we replace the readLine( ) method with a read( ) method to read the name of the file or the pathname. If we do not read sufficient number of characters, then the name of the file stored in the String object *filename* would be incorrect and this could be detected through the current implementation of the sanitize( ) method (Fig. 3) itself, as the last four characters of the file has to end in ".txt". In the context of lines 62 through 71 (boxed in Fig. 6), there would not be a problem in reading certain number of characters (rather than a line of characters) for every read operation, because—whatever is read is stored as a String and is sent across the socket. In order to preserve the text structure, we have to simply use the print( ) method instead of the println( ) method of the PrintStream class. If there is a line break in the text of the file, it would be captured through an embedded '\n' line break character and sent across the socket.

In this section, we choose to read 20 characters for each read operation at both the instances and replace the readLine( ) method with the read( ) method accordingly. In



```
48          BufferedReader brSocketInput = new BufferedReader(new InputStreamReader(clientSocket.getInputStream()));
49
50          char[] filenameChars = new char[20];
51
52          int readSocketSuccess = brSocketInput.read(filenameChars, 0, 20);
53
54          if (readSocketSuccess == -1){
55             System.out.println("Error in reading from the socket...");
56             return;
57          }
58
59          String filename = new String(filenameChars);
60
61          brSocketInput.close();
62
63          if ( sanitize(filename) == 0){
64
65                  FileReader fr = new FileReader(filename);
66                  BufferedReader brFile = new BufferedReader(fr);
67                  PrintStream socketOutput = new PrintStream(clientSocket.getOutputStream());
68
69                  char[] charsRead = new char[20];
70                  int readFileSuccess = brFile.read(charsRead, 0, 20);
71
72                  while ( readFileSuccess != -1){
73
74                    socketOutput.print(new String(charsRead));
75                    readFileSuccess = brFile.read(charsRead, 0, 20);
76
77                  }
78
79                  brFile.close();
80                  fr.close();
81                  socketOutput.flush( );
82
83          }
```

Fig. 6 Modified code for the file reader server socket program to remove the denial of service vulnerability by replacing the readLine( ) method with the read( ) method (fileReaderServer_4.java)

the second case, we read every 20 characters from the file, and the last read operation may read less than 20 characters if there are not sufficient characters. The subsequent read will return −1 if no character is read. Our logic (as shown in lines 63–71 of the modified server socket program in Fig. 6) is to check for the return value of the read operation every time and exit the while loop if the return value is −1, indicating the entire file has been read. Note that the length 20 we used here is arbitrary, and could be even set to 100. The bottom line is there should be a definite upper bound on the number of characters that can be read into the character buffer, and one should not be allowed to read an arbitrary number of characters with no defined upper limit.

3 Conclusions

In this paper, we have discussed the use of an automated tool called the Source Code Analyzer (SCA), developed by Fortify, Inc., and illustrated the use of its command line and graphical user interface (Audit Workbench) options to present and analyze the vulnerabilities identified in a software program. We presented an exhaustive case study of a file reader server socket program, developed in Java, which looks fine at the outset; but is analyzed to contain critical vulnerabilities that could have serious impacts when exploited. The four different vulnerabilities we have studied in this research are: Resource Injection vulnerability, Path Manipulation vulnerability, System Information Leak vulnerability and Denial of Service vulnerability. We discussed the reasons these vulnerabilities appeared in the code and how they could be exploited if left unattended and the consequences of an attack. We have provided detailed solutions to efficiently and effectively remove each of these vulnerabilities, presented the appropriate code snippets and the results of source code analysis when the vulnerabilities are fixed one after the other. The tradeoffs incurred due to the incorporation of appropriate solutions to fix these vulnerabilities are the increase in code size and decrease in the comfort level for a naïve authentic user who could face some initial technical difficulties in getting the program to run as desired. With generic error messages that are not so detailed, an authentic (but relatively unfamiliar) user ends up spending more time to run the system as desired. The original file reader server program had 43 lines of code, and the final version of the program (fileReaderServer_5.java) contains 97 lines—thus, an increase in the size of the code by a factor of about 2.25 (i.e., 125 % increase). However, the increase in code size is worth because even if one the above 4 vulnerabilities is exploited by an attacker, it could be catastrophic for the entire network hosting the server.

Acknowledgments The work leading to this paper is funded through the U. S. National Science Foundation CCLI/TUES grant (DUE-0941959) on "Incorporating Systems Security and Software Security in Senior Projects." The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the funding agency.

References

1. Baca D (2009) Static code analysis to detect software security vulnerabilities–does experience matter?. International conference on availability, reliability and security, IEEE, Fukuoka, Japan, In, pp 804–810
2. Caseley PR, Hadley MJ (2006) Assessing the effectiveness of static code analysis. In: 1st institution of engineering and technology international conference on system safety, London, UK, pp 227–237.
3. Chess B, West J (2008) Secure programming with static analysis, 1st edn. Addison Wesley, Boston
4. Graff MG, van Wyk KR (2003) Secure coding: principles and practices, 1st edn. O'Reilly Media, Sebastopol
5. HP Fortify SCA: https://www.fortify.com/products/hpfssc/source-code-analyzer.html
6. Mantere M, Uisitalo I, Roning J (2009) Comparison of static code analysis tools. In: 3rd international conference on emerging security information, systems and technologies, Athens, Greece, pp 15–22.
7. Mcheick H, Dhiab H, Dbouk M, Mcheik R (2010) Detecting type errors and secure coding in C/C++ applications. International conference on computer systems and applications, IEEE/ACS, Hammamet, Tunisia, In, pp 1–9
8. Novak J, Krajnc A, Zontar R (2010) Taxonomy of static code analysis tools. In: 33rd international conference on information and communication technology. Electronics and microelectronics, Opatija, Canada, pp 418–422.
9. Tondel IA, Jaatun MG, Jensen J (2008) Learning from software security testing. International conference on software testing verification and validation workshop, Lillehammer, Norway, In, pp 286–294
10. Whittaker JA (2002) How to break software, 1st edn. Addison-Wesley, Boston

Document Library System Using RDF Based Inferences

Archana P. Kumar, Kumar Abhishek and Abhay Kumar

Abstract This paper makes use of Semantic web technologies for developing a model for web searching and data representation over the existing technologies. The paper here proposes the use of a novel-ontology-based semantic search engine which takes into consideration the domain or semantic meaning of the user's query. The solution adopted here takes into consideration the RDF (Resource Description Framework) based approach which helps to convert the unstructured data into structured format. Semantic Web Technology enables to capture relationship and association as metadata. This process is described by developing a website (viz. Document Library) which runs on an inference engine which suggests the users the kind of books they like based upon its reasoning. The database for such an engine has been moved away from the traditional RDBMS format and more easy-to-use XML files have been used in place. These XML files are in fact RDF/XML files, which allow inference process to be achieved in a smooth fashion.

Keywords Resource description framework · Jena · Semantic web · Inference

A. P. Kumar
Department of CSE, MIT, Manipal 576104, India
e-mail: archie.kumar321@gmail.com

K. Abhishek (✉)
Department of CSE, NIT Patna, Patna 800005, India
e-mail: kumar.abhishek@nitp.ac.in

A. Kumar
Department of IT, NIT Patna, Patna 800005, India
e-mail: abhay.kumar@nitp.ac.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_62,

1 Introduction

Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers). [1] The traditional keyword-based information retrieval technique performs keyword searching in documents by matching the keywords that users specify in their queries. The main problem with these systems is that they do not have the ability to understand the meanings of the keywords i.e. semantics. Furthermore, different documents containing same information may be represented differently which makes it more difficult to understand semantics of the keywords. Synonym and Polysemy are two prominent issues. A synonym is a word that means the same as another word, example—author and writer. A polysemy is word with multiple related meanings; for example: fan could be an electrical device or person who is some sports fan. In semantic-based information retrieval techniques, searching is performed by interpreting the meanings of the keywords. The systems, using the technique have higher results than the systems using the traditional approach. Domain ontology's are used as knowledge based to understand the meanings of the concepts. Ontology is a formal explicit specification of a shared conceptualization. Ontology is arranged in a lattice or taxonomy of concepts in classes and subclasses. Ontology together with a set of concrete instances (or individuals) of the class constitutes a knowledge base. The semantics of the keywords are identified through the relationships or associations between keywords by performing semantic similarity on them. Currently the traditional Search technique works on keyword matching and thus lacks the ability of understanding the meaning of the user's query. The scenario of such a technique fails to have the search based on the semantics of the user's query. Moreover since the search has to be carried out among the web documents, there can be a situation where the keyword may be matched in two or more web documents which might have been used in different context. Requirement of an intelligent search engine in missing which on its own interprets the user's query and finds the web documents in the same context in which the user expects.

2 Resource Description Framework

Resource Description Framework was originally developed as a metadata model. The first specification of RDF came in the year 1999 authored by Ora Lassila and Ralph Swick [2]. RDF is based on XML and is defined as a language used for representing information about resources in the World Wide Web that is adding metadata to the web resource [2]. The basic idea of RDF is to identify things (resource) using Web Identifiers. These web identifiers are known as Uniform Resource Indicator (URI). RDF also describes resources in terms of properties and properties value [2]. RDF also known as TRIPLES where triple is a magic number that is three piece of information needed to fully define a single bit of knowledge. The three pieces of information are subject, property type and property value. For example I (subject) have a name (property), which is Kumar Abhishek (property value) [3].

In English grammar rule, a complete sentence (or statement) contains two things a subject and a predicate; the subject is who or what of the sentence and the predicate provides information about the subject [3]. For example: The title of this article is “Pranab Roy”. In the above example subject is the article, and the predicate is title, with a matching value of “Pranab Roy” [3].

If the above English statement is translated to an RDF triple, the subject is the thing being described in RDF terms, a resource identified by a URI and the predicate is a property type of the resource, such as an attribute, a relationship, or a characteristic [3]. Apart from subject and predicate the specification also introduces a third component, the object [3]. In RDF, the object is equivalent to the value of the resource property type for the specific subject [3]. RDF core committee decided to represent the data model in RDF using directed label graph [3]. The RDF directed graph consists of a set of nodes connected by arcs, forming a pattern of node-arc-node. The nodes come in three varieties: uriref, blank nodes, and literals [3]. The uriref consists of a Uniform Resource Identifier (URI) reference that provides a specific identifier unique to the node. Blank nodes are nodes that don’t have a URI. The literals consist of three parts, a character string a, an optional language tag and data type. Literal values represent RDF objects only, never subjects or predicates. In RDF literals are represented by drawing rectangles around them [3].

Figure 1 depicts a graph representation of RDF statements [4]. In the figure above, the object is a string: “Uche Ogbuji”. Then the object is termed as literal in RDF, but an object could also be a resource [4].

Figure 2 depicts combination of several RDF statements into a single diagram. The expansion of RDF is done on this basis. RDF describes a Web-based resource by defining a directed graph of statements. In the above figure, the object which is a literal “Uche Ogbuji” is replaced by a URI indicating this person, which in turn is the subject of several more statements. This type of collection of RDF statements

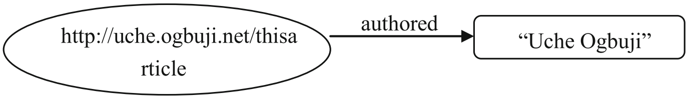


Fig. 1 RDF statement [4]

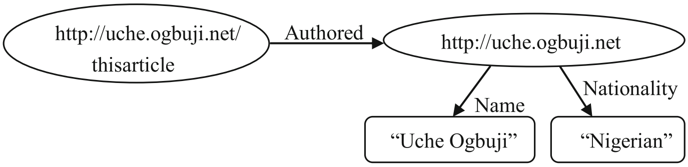


Fig. 2 A small RDF model [5]

is termed as model in RDF [4].Uniform Resource Locator are used to point to Web Documents that describe the exact meaning (semantics) of each edge type.

3 Inference Engine

Inference engine or a semantic reasoner is a software tool which is used to derive new knowledge from already existing knowledge by using rules, called as inference rules. The inference rules control all the steps for inference which is developed by the inference engine [6].

Semantic web makes use of inference engine to process the knowledge available. Let us take for instance Grandfather (Tony, Mac) |: Father (Tony, John) &

Father (John, Mac)

There are two related to the above consideration:

- Tony is father of John
- John is father of Mac

The rule devised here is—“when we find a new relationship where Tony is a father of John and then for second consideration, that John there is a Mac for which Jack is the father of Mac, then new knowledge is Tony is a grandfather of Mac” [6].

Ontology language is used to specify inference rule. While writing inference many semantic reasoner makes use of first-order predicates and based on this there two basic types of inferencing

1. Backward Chaining
2. Forward Chaining

The Backward Chaining method tries to achieve the goal by working backward that is it tries to prove a goal by finding out the truth of its condition [6]. Let us take an example of rule “if A and B then C”, the backward inference will prove C by first proving C and then proving B [6]. The Forward Chaining method is also known as data-directed inference, i.e. data gets itself in working memory. When the data is put

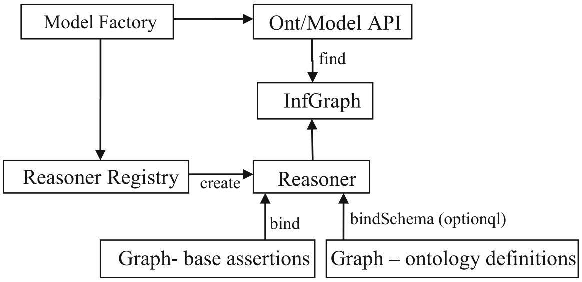


Fig. 3 Inference machinery [7]

in the working memory this will trigger rules whose condition should match new data. The rules will perform actions which may result in adding new data in memory which will trigger more [8] (Fig. 3).

4 System Design

The main aim of the paper was built a simple document library system using RDF for data representation and Storage and then implying RDF based inferences to suggest user for books. The goal was to build a system which moves out of traditional RDBMS. The implementation was done in java using an API named JENA which is used to implement RDF in JAVA.

The workflow architecture of the system makes use of the model-view-controller concept (MVC). There are different labels of interaction implemented within the application. A combination of HTML, JSP and JavaScript is used to display and input the data to the system. Two HTTP Servlets are used to control the information flow and a class within the package semantic search called SemanticSearch is used to be a kind of Java Bean and model because it capsules the Jena Functionality in points of accessing the owned private RDF graph model. The following classes were used in building the application.

(A) Schemagen

The Schemagen class within this project is a subclass of the jena.schemagen and is implemented to forward configurable running parameters to the jena class. It provides the functionality of transforming RDF Schema to an equivalent Java vocabulary class usable within Eclipse by configurate the running parameter.

(B) Vocabulary: VBOOK , VCD & VUSER

The vocabulary files are created with the name as VBOOK, VCD and VUSER which will be used by Jena. VBOOK expresses a book entity with attributes like title, author, amount, etc. VCD is built as an extension of a book by being added as an enclosure to a book. By default it is always possible to use every property described by a vocabulary class several times. That means the programmer has to take care about attribute occurrence if that should be limited. VUSER is used to express a user entity with the required properties of name, email, occupation, user name and password.

(C) Controller: Search & Admin

The controller Search and Admin are two different servlets. If a normal user searches for a book he will always send and receive his data from the Search Servlet that uses the SemanticSearch Bean to get its data but does no modification on the RDF data. Its functions are to check if a user is logged in or not and to provide a transformation of the delivered data from the SemanticSearch Bean because the bean only delivers XML or RDF data. XSL files are to execute a XSL files are used to execute a XSL transformation throw JavaX functionality (provided by the the-javax.xml.transform package) to produce the displayed HTML code. In this case XSLT is not used because it is executed from the servlet on the server. This hides every RDF and XML data from the user of the system and makes it possible to deliver more common data from the SemanticSearch Bean without taking care about possible security problems because of displaying all data about a book to the user. This is important because XSLT display its xml data source in the source code of the html page even if not all of them are displayed in the HTML representation. The Admin Servlet is responsible for using the SemanticSearch Bean to update and delete data from the RDF data file. It maps request parameters depending on the selected action that is controlled by a submitted action parameter.

(D) Bean : SemanticSerach

In the sense of reusability the SemanticSearch bean is a Java class that offers various functionalities to handle RDF data through Jena API and ARQ. It owns a private Model that stores the RDF data and offers functions to load and save this data to a file by a given filename. The delete and property change procedures are common and able to handle each kind of vocabulary not only of VBook type. Adding resources is limited on books but could be extended to other vocabularies with only some modifications. There is one more file developed named testjena.java testjena.java file which is used to set up and built the inference engine and also to run the engine [9].

5 Key Features of the System

The main highlight of this application can be listed as follows:

5.1 Data Representation in RDF

RDF based approach which helps to convert the unstructured data into structured format. This is because RDF is used to represent information modeled as a graph—a set of individual objects along with a set of connections among those objects. RDF represents the graph as triples as in subject, property and object. In the following code the subject is http://protege.stanford.edu/kb#Basics of DotNet object is Smith with the predicate author.

```
<rdf:Description
rdf:about="http://protege.stanford.edu/kb#Basics of Dot
Net">
    <kb:Count>3</kb:Count>
    <kb:Price>89.23</kb:Price>
    <kb:Currency>$</kb:Currency>
    <kb:Author>Smith</kb:Author>
    <kb:Title>Basics of DotNet</kb:Title>
</rdf:Description>
```

5.2 Searching of Data is Done Through a New Query Language SPARQL

Resource Description language is a graph based language, for querying RDF a new graph-based query language is used known as SPARQL. This application makes use of SPARQL for query and retrieving

Some benefits of SPARQL include:

1. Queries RDF data. If your data is in RDF, then SPARQL can query it natively.
2. Implicit join syntax. SPARQL queries RDF graphs, which consist of various triples expressing binary relations between resources, by specifying a sub-graph with certain resources replaced by variables. Because all relationships are of a fixed size and data lives in a single graph, SPARQL does not require explicit joins that specify the relationship between differently structured data. That is, SPARQL is a query language for pattern matching against RDF graphs, and the queries themselves look and act like RDF.
3. SPARQL is often an appropriate query language for querying disparate data sources (not sharing a single native representation) in a single query. Because RDF represents all data as a collection of simple binary relations, most data can be easily mapped to RDF and then queried and joined using SPARQL.

Steps below demonstrate querying using SPQRQL in JAVA

1. Build query string. Username and Password identifies and authenticate a user.

```
String queryString1="PREFIX kb:<"+VUSER.getURI()+">"
       +"\n"
       +"SELECT ?interest "
       +"WHERE {?x kb:Uname ?username. ?x kb:Password
           ?password. ?x kb:Interest ?interest. "
       +"FILTER (regex(str(?username), $\"" +
 "^"+userName
       + "\")"
       +" && regex(str(?password), \"" + "^"+password +
      "\"))}";
```

2. Create Query and QueryExecution object

```
Query query=QueryFactory.create(queryString1);
QueryExecution qexec = QueryExecutionFactory.create(query
                                                ,model);
```

3. Execute query and fetch result by iterating through ResultSet obtained

```
ResultSet results1 = qexec.execSelect();
QuerySolution soln= results1.nextSolution();
```

4. Get the Interest Suggestion

```
RDFNode userint=soln.get("?interest");
```

5.3 *Auto Recommendation of Books as per Profile During Sign Up*

The user.rdf file consists of the different users who have registered their names with their interest in this application. This file also contains their usernames and passwords to be matched when they visit the application again and also helps in providing the references to other recommended books [10]. This is done by using an inference engine which helps to recommend books as per the user’s interest.
We used steps similar to mentioned in Sect. 5.2 to get the recommendation with query string modified as below

```
String queryString1=="PREFIX kb:<"+VUSER.getURI()+">"
        +"\n"
        +"SELECT ?recommendation ?username ?password "
        +"WHERE {?x kb:Uname ?username. ?x kb:Password
        ?password. ?x kb:recommendation ?recommendation.
"
        +"FILTER (regex(str(?username), \"" +
"^"+userName
        + "\")"
        +" && regex(str(?password), \"" + "^"+password +
        "\"))}";
```

The inference is implemented using JENA inference rule [11–13]. The inference in the document library employed based on the user input for searching book for example if a user selects a book on XML he will automatic get recommendation for books on JAVA and .NET as XML can be implemented using JAVA as well as in .NET enviornment. The rules were written as follows:

```
@prefix kb: http://protege.stanford.edu/kb# @prefix xs:
http://www.w3.org/2001/XMLSchema#
[likec++: (?s kb:recommendation "Like Cpp"^^xs:string) <-
(?s kb:Interest ?i) regex(?i,’C’)]

@prefix kb: http://protege.stanford.edu/kb# @prefix xs:
http://www.w3.org/2001/XMLSchema#
[likejava: (?s kb:recommendation "Like Java"^^xs:string)
<-(?s kb:Interest ?i) regex(?i,’Cpp’)]
```

5.4 Indication of Books on a Rule of Count

Here based on number of books i.e. if the number of books in the database is less than 5 then the color of the title of the book will appear red in the semantic search engine.

```
<xsl:when test="wt:binding[@name=’count’]/wt:literal
&lt;=5">
<a style="color:red">
```

Similarly we have implemented if the number of books in the database is less than 50 then the color would be indicated in blue else the color will be black.

```
<xsl:when test="wt:binding[@name=’count’]/wt:literal
&lt;= 50">
<a style="color:blue">
```

5.5 XML Transformations Done to Reflect Special Characters ’+’, ’.’ and ’#’

Here if the name of the book has ‘+’ sign i.e for C++ then in the database it will be replaced with ‘p’ so in the database the book will be stored with the title as ‘Cpp’. But when the book is going to be shown in the search engine during the entire process the book will be reflected as ‘C++’. Same approach can be used for C#, .Net where ‘C#’ will be stored as ‘CSharp’ and ‘.Net’ as ‘DotNet’.

```
<xsl:when test="contains(\$tit,’Cpp’)">
<xsl:copy>
<xsl:value-of select="substring-before(\$tit,’Cpp’)"/>
<xsl:value-of select="’C++’"/>
<xsl:value-of select="substring-after(\$tit,’Cpp’)"/>
</xsl:copy>
</xsl:when>
```

6 Conclusion

The paper converts the unstructured data structure into a structured format which is done by the RDF. There is very much possibility where one could argue that RDF is just another simple XML format. But that is not true at all because they just look similar in its XML representation. This is because RDF is a subtype of XML but the difference could be found in the data model. XML in common is a tree structure whereas RDF is a graph. That is the main difference and may be an advantage if one assumes that there are more effective algorithms to deal with graphs then with trees.

This project has been a realization of a different approach along with an entirely different framework using Jena, Protégé and Eclipse. Much theory is put in the area of Semantic Web, but this project is an attempt to realize a Semantic Web Application wherein inferences have come to materialize. An exposure to a variety of SPARQL queries and a good learning of IDE Eclipse has made this project to realize a major deal of theoretic approach.

References

1. Manning CD, Raghavan P, Schütze H (2008) Introduction to information retrieval. CambridgeUniversity Press, New York
2. http://www.w3.org/TR/rdf-primer
3. Shelley Powers (2003) Pratical RDF. O'Reilly, Sebastopol
4. Gruber TR (1993) A translation approach to portable ontology specification. Knowl Acquis 5(2):199–220
5. Olgo G (2004) Inference Engines-Semantic Web Cross Up Project
6. RDF Primer: http://www.w3.org/TR/2004/REC-rdf-primer-20040210/
7. IBM http://www.ibm.com/developerworks/library/w-rdf
8. http://planetrdf.com/guide/
9. http://whatisasynonyms.com/
10. Chaoqing Lv, Kobayashi T, Agusa K, Wu K, Zhu Q (2009) Image semantic search engine. First international workshop on database technology and applications
11. JENA http://jena.sourceforge.net/tutorial/RDF_API/index.html
12. http://jena.sourceforge.net/inference/rules
13. http://hydrogen.informatik.tu-cottbus.de/wiki/index.php/JenaRules

Part VIII
Workshops: The Fourth International workshop on Wireless and Mobile Networks (WiMoNe-2012)

Spread and Erase: Efficient Routing Algorithm Based on Anti-Message Info Relay Hubs for Delay Tolerant Networks

Shikha Jain and Sandhya Aneja

Abstract Absence of end-to-end connectivity of nodes in Delay Tolerant Networks (DTNs) posses a challenge towards designing an efficient routing protocol for these networks. In DTNs with nodes having unlimited buffer sizes, Epidemic routing [16] provides an optimal solution in terms of message delivery ratio and message latency but it suffers with disadvantage of large message overhead ratio. Epidemic routing with Vaccination system (EVS) proposed by [18] improves the problem of message overhead ratio in epidemic routing by immunizing the system from duplicate message copies. We compared EVS with other routing techniques in terms of message overhead where multiple sources communicating with a single destination and observed that it performs better. In this paper we propose a routing algorithm that further minimizes the message overhead ratio of EVS and improves in both message delivery ratio and message latency. In DTNs, special nodes called Message Ferry and Data Mule have been proposed to carry data from a source node to a destination node. We propose in this paper special node called anti-message Info relay hub to erase duplicate message copies instead of routing. We compare our algorithm with EVS and social grouping based algorithm [2] for social as well as non-social scenarios and show using simulation results that our routing algorithm 'Spread and Erase with anti-message Info relay hubs' outperforms in both scenarios.

S. Jain (✉) · S. Aneja
Institute of Informatics and Communication,University of Delhi South Campus,Delhi110021, India
e-mail: shikhaa_88@yahoo.com

S. Aneja (✉)
e-mail: skhurana@cs.du.ac.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_63,

1 Introduction

Delay Tolerant Networks (DTNs) have been proposed as highly challenged networks in which end-to-end connectivity between the nodes is not always present. Examples of such type of networks include terrestrial networks connecting mobile wireless devices, outer-space networks etc. To overcome intermittent connectivity of sparsely placed nodes of these networks two approaches have been used till now (i) data is transferred using special nodes e.g. Kaash [1] and DakNet [9]. Kaash gathers data from nodes using physical device transport and DakNet suggested data transfer among nodes by copying it to removable drives and allows to physically carry the drives over vehicles, (ii) New routing algorithms [2, 6, 7, 16] are designed. Opportunistic routing is one of basic routing strategy used according to requirements of scenarios of DTNs. This routing strategy includes epidemic routing based algorithms, past encountered history based routing algorithms and social grouping based routing algorithms. The main objective of using these routing protocols is to increase the delivery ratio while decreasing the resource consumption and latency.

Social grouping based routing algorithms are being proposed for efficient delivery of data packets in comparison to epidemic. Most of researchers compare their protocol with epidemic protocol that explores all possible paths and sends packets to the explored paths with the hope that one of the paths will deliver the packets to the destination. We show that comparing social based algorithms with basic epidemic routing algorithm is not correct. We believe that there is a need to improve basic epidemic algorithm instead of using social patterns for routing in delay tolerant networks. The problem with epidemic routing protocol is duplicate messages due to its inherent nature of sending packets to all paths available in network. Zhang et al. [18] proposed epidemic routing protocol with immunity and vaccination systems to reduce duplicate messages in the network. Immunity based epidemic routing protocol disseminate information in reverse direction once messages are delivered to their respective destinations. In epidemic routing with vaccination system(henceforth referred as EVS), this information is also passed to nodes which have not yet received the packet but in future may receive it. Padma [8] compared immunity based routing protocol with basic epidemic routing protocol and shown that immunity based protocol performs better than the basic epidemic routing protocol in terms of both in delivery ratio and overhead ratio.

Social grouping based routing protocol, for example, Dynamic Social Grouping (*DSG*) [2] as proposed by Roy et al. uses social patterns of nodes to increase the efficiency of routing in Delay Tolerant Networks. Roy compared DSG with basic epidemic and probabilistic routing method and shown that DSG protocol performed (i) better than probability routing and similar to epidemic routing for message delivery ratio and message delivery time, and (ii) better than both probability routing and epidemic routing in terms of overhead. Therefore, Roy's claimed that social based routing algorithm i.e. DSG performs better than epidemic routing in terms of overhead.

We compared DSG with EVS instead of basic epidemic routing. We observed EVS performs better than DSG in terms of message delivery ratio, average message latency, and overheads. Therefore, future research should compare newly designed protocols with EVS, as proposed by Zhang. Further, we modified EVS, in order to remove duplicate messages faster. We show in this research that spread and erase, specialized node based algorithm is more efficient than basic EVS and performs better than social based routing protocols also.

In social scenario like rural area with intermittent connectivity, social context based algorithms are required to route the information efficiently but in other DTN scenarios like under-water, outer-space and terrestrial networks these algorithms may not perform better. We need routing algorithm that is robust in terms of scenarios (social or non-social). We propose a robust approach for social and as well as for non-social scenarios. In our approach, we tried to optimize EVS in terms of communication overhead using specialized nodes in the network. In social scenarios like rural area of intermittent connectivity as well as non-social scenarios, the assumption to have special node is realistic as in we generally have special nodes in rural area to support networking. The algorithm has two phases. The first phase, Spread, is similar to epidemic algorithm but once message is delivered to destination, it starts its second phase, Erase, by initializing a list with identifiers (IDs) of delivered messages. Erase phase works like EVS. To immune the system from message overhead each node at the time of contact exchange their information about messages which are delivered to destination. Specialized non-intelligent high range nodes are taken in network to further increase the speed of vaccination. These nodes just immune the system and do not participate in message forwarding and buffering. We found that even in a network of 85 nodes simultaneously communicating to destination a few specialized nodes were sufficient to achieve the delivery probability higher than EVS and DSG. Message overhead of our protocol was much less than aforementioned schemes.

2 Related Work

Vahdat et al. [16] presented the epidemic routing algorithm for delay tolerant networks. They proposed that each node maintain a summary vector that comprises *IDs* of messages present in its buffer. When two nodes meet they exchange their summary vectors. This enables the nodes to identify the messages which are not present in their buffer and they request for them. In order to reduce communication overhead, hop count field is used to limit the number of hops that each message can traverse. Many researchers presented other algorithms to further minimize communication overhead of epidemic routing. We categorize the work on improvement of epidemic algorithm into three different approaches. First approach [13, 14, 17] focuses on controlling message copies by relaying only a limited number of message copies in the network. This approach improves communication overhead but suffers from long delays as message copies are limited and delivery of message depends on when node having complete message comes in contact with destination. Second approach [8, 18] focuses on erasing the message copies in the nodes' buffer after these messages are delivered. Authors [8] claimed to achieve 15 % increase in delivery ratio and similar delay performance at less than half the buffer capacity than epidemic algorithm. [18] presented an ordinary differential equation(ODE) based model for epidemic and modified epidemic algorithms called immunity and vaccination protocols. Third approach [2, 3, 5, 7] uses history of encounters instead of opportunistic forwarding like epidemic in order to limit forwarding of message copies. In ProPHet [7] history of encounters is used to compute predictability to deliver the packet to a destination. A node with lower delivery predictability when comes in contact with another node, it transfers a message to it if it has higher delivery of predictability. Authors used community—based scenario to save communication overhead in comparison to epidemic algorithm. Social group based routing further extends probabilistic routing to improve message delivery with constrained buffer sizes and communication overhead over epidemic algorithm. These algorithms capture social groups knowledge in addition to history of encounters to forward the message to next node. Simbet [3] assumes existence of social groups and uses centrality and similarity measure for each node in communities to increase the probability to deliver a message. Bubble rap [5] extends their work and provides clustering and K-clique method to create social groups. Authors proposed routing based on assumption that each node within a social group has two labels local and global. A packet is routed through a node if it matches first with local label and then with global label. The social groups used by these algorithms are static i.e. the groups once formed do not get updated as a new node joins or deletes itself from group. In [2], authors proposed a routing algorithm which forms dynamic social groups using regular contact pattern of nodes. Each node maintains individual probability as a measure of number of messages forwarded by it and also maintains group probabilities that measures degree of messages forwarded by group and its contact strength in group. Messages are routed through nodes that have higher probability to deliver a message to the destination. Authors claimed to achieve message delivery ratio as epidemic algorithm with much lower communication overhead.

3 Problem Statement

Seeing the advantages associated with EVS, we pose the problem in DTNs with EVS as routing protocol to further minimize its communication overhead. In DTNs various type of specialized nodes referred as message ferry [4, 15, 20], Infostation [12] and datamule [11] have been proposed to improve the message delivery in the network. These nodes are used to collect the data from various nodes and to transfer data to destination. Many algorithms are proposed [19] on trajectory and scheduling of contacts of these nodes with other nodes to meet the requirements of high traffic load of large network scenarios. We propose to use special nodes (called Anti-message Relay Hubs) for relaying the anti-message information to the forwarding nodes instead of collecting data in DTNs. We have assumed that destination called base station is fixed in the network and various nodes are trying to send data to base-station simultaneously.

4 Algorithm

We propose spread and erase routing scheme with anti-message Info relay hubs as a modification of epidemic algorithm to improve over message overhead. Motivation of our idea (spread and erase) is that the increase in speed of vaccination and immunization system in epidemic routing will improve significantly over the message delivery ratio, average message latency as well as message communication overhead.

Spread and Erase is a two phase algorithm. In spread phase each node spreads messages present in its buffer to the other nodes that comes in its contact. In erase phase like vaccination system (*VS*) it erases delivered messages from the nodes buffer. Both phases execute simultaneously. During erase phase, we propose specialized non-intelligent nodes to increase speed of vaccination in the network called anti-message Info relay hubs. These nodes move randomly in the network to gather information of delivered messages. Hubs are high range non-intelligent nodes in comparison with other nodes and thus do not forward data packets. Because of high range, these nodes are in contact with each other and a significant number of other nodes. Hubs have two interfaces for communication for example Bluetooth and Wifi. One interface Wifi it uses for communicating with other hubs and other Bluetooth it uses for communication with normal node of network. Hubs help to erase already delivered messages (spreaded messages) from buffer of ordinary nodes. Since, these nodes do not forward data messages so they do not need bundle layer to store messages. We propose that the increase in speed and number of hubs will further improve the message overhead ratio of the algorithm. We verified this fact using experimental work.

For a *DTN* with *n* nodes using Spread and Erase scheme if a node wants to communicate with other node, it creates a message and keeps in its buffer. Each node (excluding hub nodes) maintains a list which consists of *IDs* of messages in its buffer. This list is known as summary vector (*SV*). According to algorithm, each node (including hub nodes) also maintain one more list called acknowledgment summary vector (ASV) which consists of IDs of messages delivered to destination. The list of ASV is maintained as a queue. Its size can be fixed by taking a value sufficiently large in comparison to maximum hop count traversed by the message. When two nodes come in contact with each other they first exchange their $ASVs$ and SVs. Using SV both node exchange messages to each other which are not present in other nodes' buffer. Also nodes delete messages which are delivered to destination and is present in buffer using ASV. They synchronize their $ASVs$ so that in future if a message which is delivered to destination is received it is discarded. When a node come in contact with hub node, they also exchange their $ASVs$. Node deletes the messages from its buffer which are already delivered to destination using received ASV from hub node but the hub node just updates its ASV.

5 Simulation Study

5.1 Simulation Scenario and Data Sets

We simulated our protocol using Opportunistic Network Environment (*ONE*) Simulator. Protocol was run for one set of real data traces called Haggle data set and one inbuilt dataset of *ONE* simulator. These data sets were taken for one social scenario with varying communities called Infocom05 [10], and one with general scenario of Helsinki city. In Infocom05 [10], the devices were distributed to approximately fifty students attending Conference IEEE Infocom in Grand Hyatt Miami for four days from different countries and origin.

We ran the simulations for 50,000 s. The total number of messages generated in the network was 2271. The number of node was varied from 85–41 depending on the respective data sets. Message size was taken 512 Kb to 1 MB . Message TTL (Time to live) was set to 1200 min. Messages were generated randomly by nodes in every 20–25 s interval for a fixed destination node with ID 0. Message Delivery Ratio (MDR), Average Message Latency (AML), Message Overhead Ratio (MOR) and Routing Overhead (RO) metrics were used to compare three DTN routing protocols: EVS [18], DSG [2], and our protocol Spread and Erase (SE). MDR is the ratio of the number of delivered messages to the number of messages created and AML is the average delay between sending the message by the source and its receipt at the destination. This includes all the delays caused during route acquisition, buffering and processing at intermediate nodes, etc. RO is the number of control messages generated for total number of data messages delivered and MOR is ratio of subtraction of messages relayed and messages delivered by total number of messages generated.

5.2 Simulation Results

We observed that for non-social data, EVS and Spread and Erase algorithms outperformed in terms of message delivery ratio, message overhead, average buffer time and routing overhead over DSG (see Fig. 1). Spread and Erase improved message overhead and average buffer time in comparison EVS with almost same message delivery ratio. For social data set also, EVS and Spread and Erase improved in terms of message delivery ratio, message overhead, average buffer time and routing overhead over DSG (see Fig. 2). We showed in Fig. 3 performance of Spread and Erase with varying number of hub nodes. We observed that message delivery ratio does not change with number of nodes but shows a linear improvement in message overhead. We observed the same behavior by varying speed of nodes (see Fig. 3).

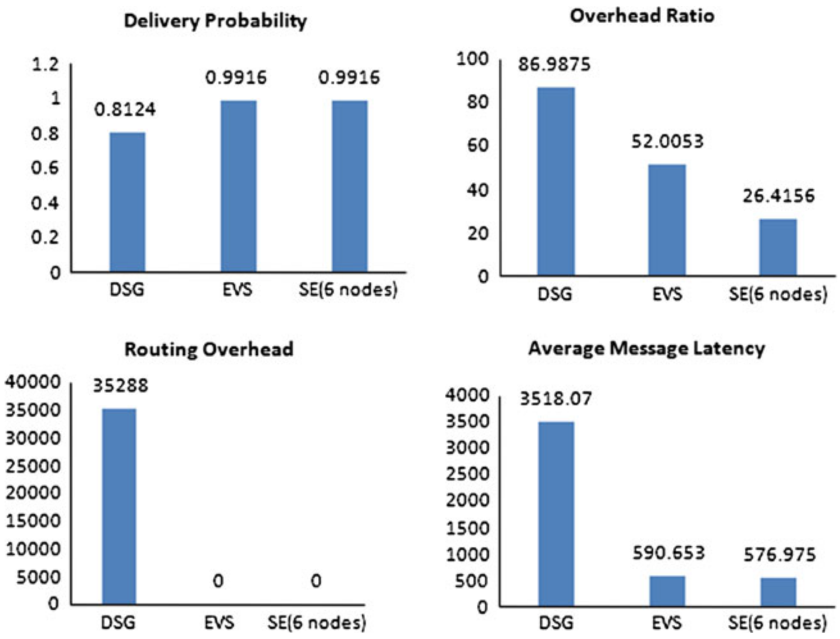


Fig. 1 Comparison for DTN inbuilt dataset

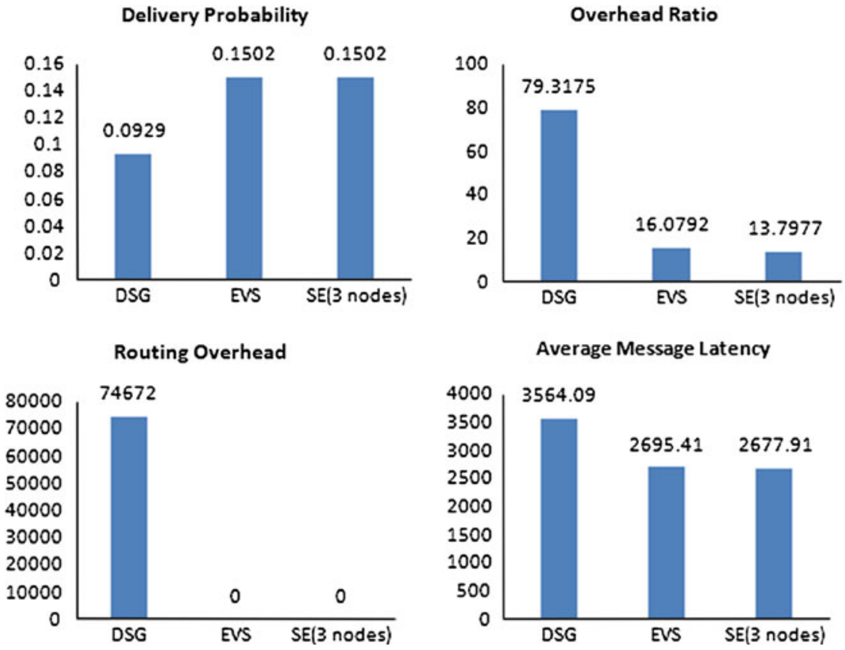


Fig. 2 Comparison for Haggle dataset

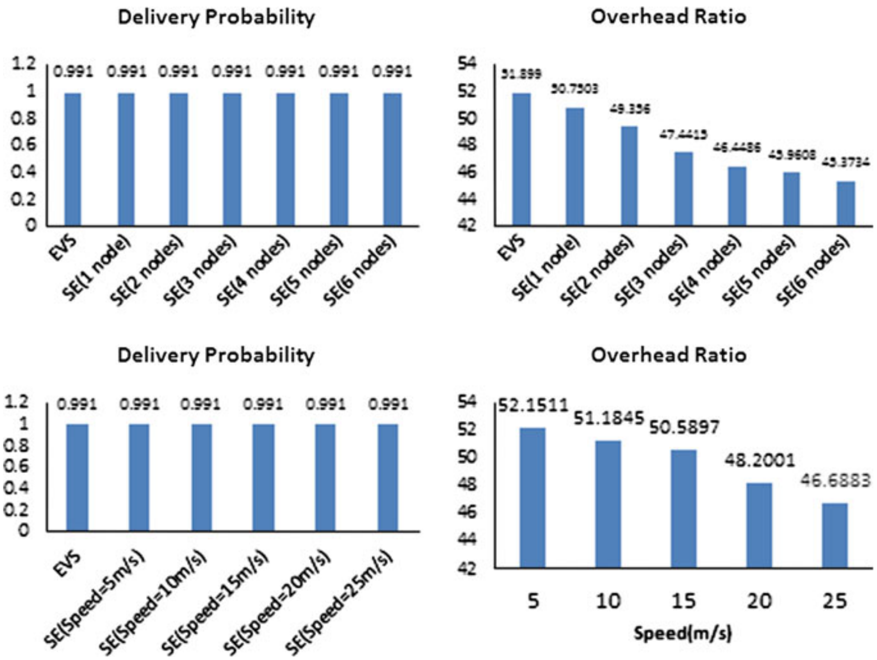


Fig. 3 Comparison for DTN inbuilt dataset by varying number and speed of nodes

6 Conclusion and Future Work

We used anti-message relay hubs to erase the message duplicate copies in the network. We observed that a few hubs with speed comparable to a car, motor bike are sufficient for a network of large size. For non-social as well as social scenario, we observed that even randomly moving mobile hubs in the network immune the nodes at sufficiently high speeds. In future to further reduce the number of hubs, we propose to work on scheduling and finding the trajectory for hubs.

References

1. Anantraman VKV, Mikkelsen T, Ohno L (2002) Handheld computers for rural healthcare, experiences in a large scale implementation. In: Proceedings of development by design
2. Cabaniss R, Madria S, Rush G, Trotta A, Vulli SS (2010) Dynamic social grouping based routing in a mobile ad-hoc network. In: IEEE international conference on mobile data management
3. Daly EM, Haahr M (2007) Social network analysis for routing in disconnected delay-tolerant manets. In: Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing, MobiHoc '07, pp 32–40
4. Gu Y, Bozdag D, Ekici E, Ozguner F, Lee CG (2005) Partitioning based mobile element scheduling in wireless sensor networks. In: Proceedings of the 2nd annual IEEE conference on sensor and ad hoc communications and networks

5. Hui P, Crowcroft J, Yoneki E (2008) Bubble rap: socialbased forwarding in delay tolerant networks. In: Proceedings of MobiHoc
6. Jain S, Fall K, Patra R (2004) Routing in a delay tolerant network. SIGCOMM Comput Commun 34:145–158
7. Lindgren A, Doria A, Schelen O (2003) Probabilistic routing in intermittently connected networks. SIGMOBILE Mobile Comput Commun Rev 7:19–20
8. Mundur P, Seligman M, Lee G (2008) Epidemic routing with immunity in delay tolerant networks. In: Military communications conference, pp 1–7
9. Pentland AS, Fletcher R, Hasson A (2004) Daknet: rethinking connectivity in developing nations. Computer 37:78–83
10. Scott J, Gass R, Crowcroft J, Hui P, Diot C, Chaintreau A (2006) CRAWDAD trace http://www.cambridge/haggle/imote/infocom (v. 2006–01-31)
11. Shah RC, Roy S, Jain S, Brunette W (2003) Data mules: modeling a three-tier architecture for sparse sensor networks. In: Proceedings of the first IEEE international workshop on sensor network protocols and applications
12. Small T, Haas Z (2003) The shared wireless infostation model: a new ad hoc networking paradigm (or where there is a whale, there is a way). In: Proceedings of the 4th ACM international symposium on mobile ad hoc networking & computing, MobiHoc ’03
13. Spyropoulos T, Psounis K, Raghavendra CS (2004) Single-copy routing in intermittently connected mobile network. In: Proceedings of IEEE Secon
14. Spyropoulos T, Psounis K, Raghavendra CS (2005) Spray and wait: an efficient routing scheme for intermittently connected mobile networks. In: Proceedings of the 2005 ACM SIGCOMM workshop on delay-tolerant networking, WDTN ’05, pp 252–259
15. Tariq MMB, Ammar M, Zegura E (2006) Message ferry route design for sparse ad hoc networks with mobile nodes. In: ACM MobiHoc 06
16. Vahdat A, Becker D (2000) Epidemic routing for partially connected ad hoc networks. In: Technical report CS-200006 Duke University
17. Wang Y, Wu H (2007) Delay/fault-tolerant mobile sensor network (dftmsn): a new paradigm for pervasive information gathering. IEEE Trans Mobile Comput 6:1021–1034
18. Zhang X, Neglia G, Kurose J, Towsley D (2007) Performance modeling of epidemic routing. Comput Netw 51:2867–2891 (Elsevier)
19. Zhao W, Ammar, M, Zegura, E (2004) A message ferrying approach for data delivery in sparse mobile ad hoc networks. In: Proceedings of the 5th ACM international symposium on mobile ad hoc networking and computing, MobiHoc ’04 (2004)
20. Zhao W, Ammar M, Zegura E (2005) Controlling the mobility of multiple data transport ferries in a delay-tolerant network. In: INFOCOM

Novel Architecture of Adaptive and Optimized Policy Based Handover in MANET

Nidhi Parashar and A. K. Vatsa

Abstract MANET's random behavior and absence of any central intelligence to gather unambiguous knowledge about user contexts complexes QoS maintenance and hampers proper utilization of network resources resulting into unnecessary handovers. In such network uninterrupted connectivity requires the development of an efficient handover mechanism. A set of policies can be applied to efficiently manage the network resources. Therefore in this paper we propose an adaptive and optimized policy based handover mechanism which is based on explicitly designed policies and predicts the high time of handover need on the basis of application specific needs of individual freely roaming mobile nodes, avoiding unnecessary handovers and provides efficient handover procedure with optimized resource consumption, reduced latency and interruption time.

Keywords Mobile ad-hoc network (MANET) · Service discovery · Application specific threshold · Energy-distance factor · Load balancing · QoS · RSSI · Handover

1 Introduction

With the recent advances in the development of wireless communication technologies an Infrastructure less network MANET is quickly growing for its ability to form the network of dynamic topology without any fixed or predefined infrastructure. The mobile nodes are constrained with high mobility and energy limitations therefore a

N. Parashar (✉)
Shobhit University, Meerut, Uttar Pradesh, India
e-mail: nidhi.csit@gmail.com

A. K. Vatsa
Department of Computer Science, University of Missouri-Columbia, Columbia, MO, USA
e-mail: avimanyou@rediffmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_64,

hierarchical architecture is needed to achieve performance guarantee in a large scale MANET thus MANET nodes are grouped into clusters. Under a cluster structure resources can be spatially reused to increase the system capacity and two clusters may use the same frequency or code set if they are not neighboring clusters. Internal events in the cluster can be coordinated by its cluster head. For satisfactory performance of MANET an effective cluster formation, cluster head selection mechanism considering all the network dynamics is needed along with a handover mechanism that performs efficiently in various network conditions.

Policy based handover mechanisms have been done in past but no one specifies the design and mechanisms for policies to be considered while taking handover decision [3]. Here an adaptive and optimized handover mechanism based on certain policies has been deduced that significantly avoids the overhead created due to unwanted handover initiation. At the same time detail mechanisms for different policies are optimized in terms of network throughput and QoS parameters resulting into an optimal handover decision reducing the handover latency and processing delay. Energy distance factor [5] facilities the energy efficient decisions in different phases of the proposed mechanism.

Rest of the paper is organized as follows. Section 2 gives the literature review while proposed work is discussed in Sect. 3. Section 4 concludes the contribution of this paper along with the future scope in Sect. 5.

2 Background

Dynamic Cluster Protocol [8] having five states such as un-clustered state, orphan state, election state, cluster node state, and cluster head state are identified for a MANET. Detection and control mechanism for DDOS attacks over reputation and score based MANET [6] gives clustering technique using reputation and score value of nodes but the important parameters like distance from centroid, processing capability and mobility of a node in cluster head selection procedure have not been considered.

A distributed directory based service discovery mechanism [2] operates in a proactive mode and selects a provider based on distance and service capability of the provider. A probability based node selection method [5] identifies the intermediate node with optimum stored energy that could withstand through duration of connection. A handover scheme considering the current active application in use rather than a single fixed threshold value and dynamically changing the threshold value depending upon the currently active applications [1] have been used to avoid unnecessary handovers.

Policy Based Fast Handover Mechanism for MANET [4] provides proactive and reactive handover approaches based on policy for nodes of MANET but parameterized design for different policies in changing situations have not been explored. In addition to the more dynamic and adaptive mechanisms for cluster formation and cluster head selection, most of the policy mechanism to provide best handover decisions and reducing the overhead caused by unnecessary handovers.

3 Proposed Work

This section presents the proposed Architecture and mechanism in the Sects. 3.1 and 3.2.

3.1 Architecture of Optimized Policy Based Handover

The proposed architecture in Fig. 1 includes the following modules.
Cluster formation module. It co-ordinates cluster formation in MANET. All mobile nodes in MANET are organized into clusters on the basis of Device type, Bandwidth, RSSI value, Services required, Distance between nodes.
Cluster Head Selection module. It facilitates the score based weighted selection of a node as Cluster Head.
Policy Module. It comprises different policy mechanisms that guarantee efficient resource utilization along with better QoS inside MANET.

Load Manager manages the load balancing task inside a cluster. On the basis of current load and remaining battery power of the cluster nodes the load among them is distributed for the smooth and efficient working of cluster.

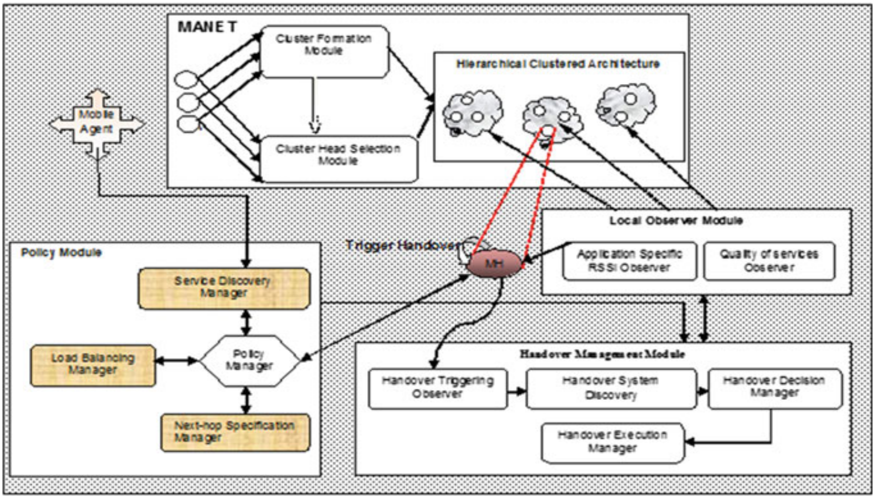

MANET
Cluster Formation Module
Cluster Head Selection Module
Hierarchical Clustered Architecture
Mobile Agent
Trigger Handover
MH
Local Observer Module
Application Specific RSSI Observer
Quality of services Observer
Policy Module
Service Discovery Manager
Load Balancing Manager
Policy Manager
Next-hop Specification Manager
Handover Management Module
Handover Triggering Observer
Handover System Discovery
Handover Decision Manager
Handover Execution Manager

Fig. 1 Adaptive and optimized policy based handover in MANET

Service Manager manages intra cluster and intra cluster service discovery to the MN while moving from one place to another.

Next-Hop Specification Manager selects optimal cluster for a MN when it receives equal RSSI value from two or more clusters.

Policy Manager co-ordinates the working of overall policy module. It collects the updated policy management information from all the managers, compiles the gathered information and gives suggestions to MN regarding handover need and Quality of services management.

Mobile Agents. These freely moving ad hoc network nodes collect the location and description of services available in different clusters and make this global service information available to the CHs of all clusters.

Local Observer module. It observes the local functioning of a cluster and assists the CH in various cluster maintenance activities.

Application Specific RSSI Observer calculates the RSSI requirement of all active MNs within a cluster according to the currently running applications in each MN. If more than one application is running simultaneously and all applications require different RSSI then this module selects the maximum of them as $\mathrm{RSSI}_{\mathrm{AST}}$.

Quality of services Observer keeps track of all the network parameters like bandwidth, RSSI, load distribution on all MNs within cluster.

Handover Management module. It provides optimal handover when MN detects the need to move seamlessly within MANET. *Handover Triggering Observer* receives handover request from MN and invokes handover system discovery manager to find potential CHs.

Handover System Discovery Manager prepares list of target CHs on the basis of RSSI required, capability of target CHs and the context for which handover need is initiated and sends this list to Handover Decision Manager for target selection.

Handover Decision Manager selects the optimal target cluster.

Handover Execution Manager coordinates the actual handover including authentication and association to the target cluster.

3.2 Mechanism of Optimized Policy Based Handover

Optimized Cluster Creation and cluster Head Selection. In this primary phase details of cluster formation and cluster head selection mechanisms are specified. *Node Admission Control* manages the entry of a new node in MANET. While a node enters MANET it is in UN Clustered state. It sends a hello message to all neighbors. After a predefined number of tries If it does not get reply before expiry of timer (t_{uc}), it enters into Orphan state (Fig. 2).

```
NodeAdmissionControl( )
{
If (t<=$t_{uc}$)
    ApplyForCluster();
 Else
 {  Ni turns Orphan.
   If(gets GetState from CH)
      ApplyForCluster ();
   Else
      CreateNewCluster}
}
```

```
ApplyForCluster()
{For (i=1; i<=n; i++)
 If((clusterSize<MaxSize)&&
    (Ni=DType)&&($BW_{Ni}$>=$BW_T$)
     JoinCluster ();
 Elseif(clusterSize>MaxSize)
     CreateNewCluster ();
  Else
      MergeCluster();
     //if no.ofnode<MinSize
}
```

```
JoinCluster( )
{//Ni is in eligibility test and Nj is the existing node
 If(($D_{NiNj}$<=$D_{THRESHOLD}$)&&($ClusterRange_{Ni}$<=$Range_{TH}$)&&
    ($RSSI_{Ni}$>=$RSSI_{TH}$)&&(ServiceRequired€ServiceOffered))
       Ni is accepted in the cluster;}
```

```
CreateNewCluster( )
{Appoint the initiating node as CH;
   If(CH does not get any message for a time period)
      CH changes its state from CH to Orphan State;
   Else
      ClusterHeadSelection();}
```

```
    MergeCluster( )
    {//NC is total no of user clusters.
      For (j=1; j<=NC-1; i++)

If($Size_{Ci}$<MinSize||$Size_{Cj}$<MinSize)&&(Ci and Cj are in range)
  If($Size_{Ci}$<=$Size_{Cj}$)
   {//CP is the processing power of node.
     If(($CP_{CH}$>=$CP_{THRESHOLD}$)&&(CHnotoverloaded)&&
         ($ServicesOffered_{Ci}$ €$ServiceOffered_{Cj}$))
          For all nodes in Ci
              ApplyForCluster;
              ClusterHeadSelection();}
     ElseIf(($CP_{CH}$>=$CP_{THRESHOLD}$)&&($Load_{CH}$<=$Capacity_{CH}$)&&
            ($ServicesOffered_{Cj}$ €ServiceOffered$_{Ci}$))
          For all nodes in $C_j$
             ApplyForCluster;
             ClusterHeadSelection();}}
```

Cluster Head Selection process considers the following scores

Reputation Score returns trust value of the node depending upon its previous record of packet forwarding capability. Each node in cluster contains an entry {N_{ID}, S_{TRUST}, U_{TRUST}}. Let S $_{TRUST}$ and U_{TRUST} are scores for correct and Incorrect packet forwarding capability $S_{TRUST}(i, j) = 1$, if i gets service from j and $U_{TRUST}(i, j) = -1$ otherwise. Trust value can be calculated as

$$tValue = \Sigma\ S_{TRUST}(i, j) + \Sigma\ U_{TRUST}(i, j) \quad (1)$$

Mobility Score provides means to deduce stability factor of mobile node $D_{i,j}$ is the distance between two nodes i and j and MD_i is Mean distance between Ni and all its neighbors STi is the stability of node i.

$$D_{i,j} = \sqrt{(Xi - Xj)^2 + (Yi - Yj)^2} \tag{2}$$

$$MD_i = \frac{1}{N}\sum_{i=1}^{N} Di, j \tag{3}$$

$$ST_i = MD_t - MD_{t-1} \tag{4}$$

Energy Distance Score finds node that consume less energy to communicate with rest of cluster nodes resulting in less battery power consumption. Every node calculates Energy-Distance factor (ED) with other node. Node having highest ED can be preferred as CH. Let E_i and D_i is residual Energy and distance of a node respectively.

$$ED = \frac{Ei * Di}{\sum Ei * Di} \qquad //0 <= ED <= 1 \tag{5}$$

Degree_Score() gives the number of neighbor nodes.Node calculates no of neighbors on the basis of received Get State messages. *DFC( )* gives the distance from cluster centroid. Let (x_c,y_c) are co-ordinates of cluster centroid and $D_{i,c}$ is the distance of a node from centroid.

```
ReputationScore( )
{ For(i=1;i<=n;i++)
    If(tValue$_{Ni}$>=0)
      Node is trustworthy;
    Else
      Node not trustworthy;
    Return(tValue);}
```

```
Degree_Score ()
{ Degree$_{Ni}$ =0
  For (i=1; i<=n; i++)
    If(Ni gets msg from Nj)
       //where j=1to n.
          Degree$_{Ni}$++;
     Return (Degree$_{Ni}$);
}
```

```
EnergyDistanceScore( )
{For(i=1;i<=n;i++)
     Calculate ED$_{Ni}$
 Return(ED$_{Ni}$);}
```

```
Mobility_Score()
{For (i=1; i<=n; i++)
Return (ST$_i$);
}
```

```
DFC( )
{For (i=1; i<=n; i++)
 {
   $D_{i,\ c} = \sqrt{(Xi - Xc)^2 + (Yi - Yc)^2}$
 }
  Return (D$_{i,\ c}$);
}
```

For each node all of these scores are assigned a weight so that summation of all weights is unity as

$$\sum_{j=1}^{n} W = 1. \tag{6}$$

Let $W_T[Ni]$, $W_S[Ni]$, $W_E[Ni]W_D[Ni]$, $W_{DFC}[Ni]$, $W_B[Ni]$ are partial weights assigned for all scores. Global weight for each node is calculated as

$$\begin{aligned} W_G[Ni]) = {} & (W_T[Ni]^* tValue_{Ni}) + (W_S[Ni]^* ST_{Ni}) + (W_E[Ni]^* Eff_{Ni}) \\ & + (W_D[Ni]^* Degree_{Ni}) + (W_B[Ni]^* Battery_{Ni}) \\ & + (W_{DFC}[Ni]^* dfc_{Ni}) \end{aligned} \tag{7}$$

```
ClusterHeadSelection()
{
  For (i=1;i<=n;i++)//for all nodes in the cluster
  {    $TV_{Ni}$      = ReputationScore(Ni);
       $ST_{Ni}$      = Mobility_Score(Ni);
       $Eff_{Ni}$     = Energy_Distance_Score(Ni);
       $Deg_{Ni}$     = Degree_Score(Ni);
       $dfc_{Ni}$     = DFC();
       $Battery_{Ni}$ = MeasuredPower(Ni);}
  For (i=1; i<=n;i++)
  {      calculate $W_G$[Ni];}
     Min_Global_Weight $W_{Gmin}$ =MIN($W_G$[Ni]………$W_G$[Nn]);
     Assign node having $W_{Gmin}$ as CH;
     //Node having second lowest $W_G$ is assigned as $CH_{Backup}$.
     If(CH fails || $Load_{CH}$>=$Capacity_{CH}$)
        $CH_{Backup}$ is selected as CH.
     If($Battery_{CH}$< $Battery_{THRESHOLD}$)
        CH sends power low signal to neighbors and turns orphan.
        Cluster Head Selection ();
}
```

Optimal Policy Based Parameters. To discover accurate decision techniques for handover process, following policies are considered.

Load Balancing forces CH to periodically collect energy and load values from all cluster nodes. Considering $Capacity_{CH}$ is the max no of nodes CH can serve and $Load_{CH}$ is the no of nodes it is serving at present, an overloaded node generates a Relaxation Request to CH. For load balancing among clusters CHs of all clusters in the network broadcasts their load values to all the neighbor CHs. Load balancing ensures balanced clusters inside the MANET and raises the need for handover when load inside cluster is non manageable providing optimal handover triggering.

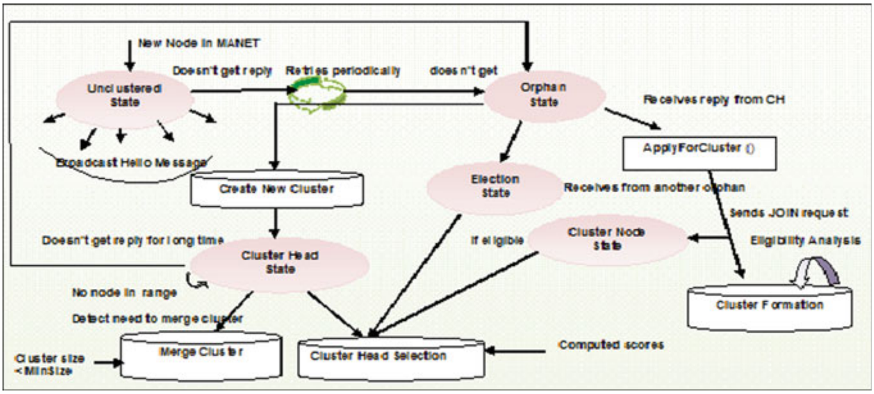


Fig. 2 Cluster creation and cluster head selection

Application Requirement results into hybrid mode service discovery given in Fig. 3 all SP nodes advertise their services to their CH proactively while SR nodes query their CH for service information whenever it is in need of any service reactively. CH is responsible for intra cluster service discovery and inter cluster service discovery is done with the help of mobile agents. Optimized service discovery mechanism reduces the risk of unnecessary handovers triggered due to requirement of services.

Next-hop Specification helps the cluster nodes who may reach to more than one CH. Let CH_{cur} current CH and CH_{alter} is alternative CH Such nodes may obtain equal RSSI value from two different CHs. Taking the current load and battery power intermediate hop count of each CH into consideration the better cluster is decided (Fig. 4).

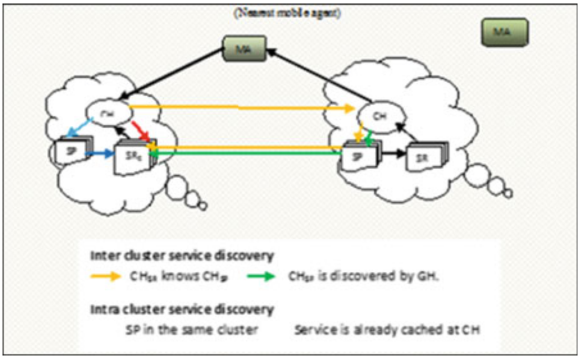


Fig. 3 Hybrid service discovery

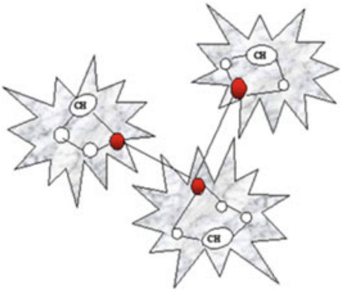


Fig. 4 Next hop selection

```
LoadBalance( )
{
  For(i=1; i<=n; i++)//for all nodes in cluster
   If(((Load$_{Ni}$>=MaxLoad)&&(E$_{Ni}$<=MinE))
       ||((Load$_{Ni}$>=MaxLoad)||(E$_{Ni}$<=MinE))
         Ni sends Relax$_{Req}$ to CH.
         CH sends address of capable nodes in Relax$_{Rep}$.
         If(Ni receives more than one address in Relax$_{Rep}$)
            For(j=1; j<=k; j++)//k possible nodes
                Compute Dij.
                MinDistance= MIN (D$_{i1}$,D$_{i2}$..D$_{iK}$);
                Nj at Minimum distance is selected
}
Service_discovery( )
{If (SP exists in same cluster)
      S$_{req}$ is forwarded to SP that replies to SR .
 Elseif(same S$_{req}$ has been cached in $CH_{SR}$)
      $CH_{SR}$ replies to SR.
 Else if(SP doesn't exist in cluster)
      If($CH_{SR}$ knows the SP)
          $CH_{SR}$ forwards S$_{req}$ to $CH_{SP}$ which replies.
      Else
          $CH_{SR}$ sends S$_{req}$ to nearest MA who forwards S$_{req}$ to $CH_{SP}$
}
NextHopSelection()
{If(MN directly connects CH$_{cur}$||MN directly connects CH$_{alter}$)
    CH at zero hop distance is selected as CH$_{target}$
 Else  //Let k intermediate nodes to CH$_{cur}$ and l nodes to CH$_{alter}$
                    ED( CH$_{alter}$) = $\frac{1}{l}\sum_{i=1}^{l}$EDi
                    ED( CH$_{cur}$) = $\frac{1}{k}\sum_{i=1}^{k}$EDi
 If(ED(CH$_{alter}$)==Max(ED(CH$_{alter}$),ED(CH$_{cur}$))&&(Load$_{CHalter}$< Load$_{CHcur}$).
    CH$_{alter}$ is selected as CH$_{target}$ }
```

Optimized Hand Over. Handover can be performed proactively or reactively according to the following stages:

Handover Triggering predicts the need for handover depending upon Load$_{CH}$, available services and RSSI$_{AST}$ value.

Handover System discovery searches for better RSSI value proactively or reactively whenever mobile node experience (RSSI< RSSI_{AST}).

Handover Decision State selects the optimal cluster. In Proactive mode to reduce the handover latency period, as soon as the mobile node gathers a number of probe advertisements it starts comparing the QoS offered in each cluster and based on various policies and user contexts selects the best cluster to get better QoS. A Prepare Handover request is sent to the target cluster so that target cluster may reserve the required resources in advance. Same policy is applied in reactive mode (Fig. 5).

Handover Execution Phase starts with $\text{Authentication}_{Req}$ sent to target CH by MN, After receiving $\text{Authentication}_{Ack}$ MN sends the Association_{Req} and waits for $\text{Association}_{\text{Ack}}$. At the end binding update informing about its new location is sent to the old CH which replies with Binding Ack.

```
Trigger_Handover()
{For (i=1; i<=n; i++)
    If(($Load_{CH}$>=$Capacity_{CH}$)||(!ServicesPresent)||($RSSI_{Ni}$<=$RSSI_{AST}$))
          Initiate System Discovery();}

Handover Decision( )
{Find CH with ($RSSI$>=$RSSI_{AST}$)&&($Load_{CHi}$<=$Load_{CHcur}$)
 For(i=1; i<=P; i++)// P is total potential CHs
   If(($ED_{CHi}$==Max($ED_{Ch1}$..$ED_{ChP}$)&&(ServicesReq€$ServiceOff_{CHi}$)&&
      ($Battery$=$Battery_{TH}$)&&($Stability_{CHi}$=Max($Stability_{CH1\ldots..CHp}$)&&
      ($Latency$<$Latency_{TH}$)
         Select CHi as target CH for handover procedure
   Else{ Repeat the process for the next CH.}}
```

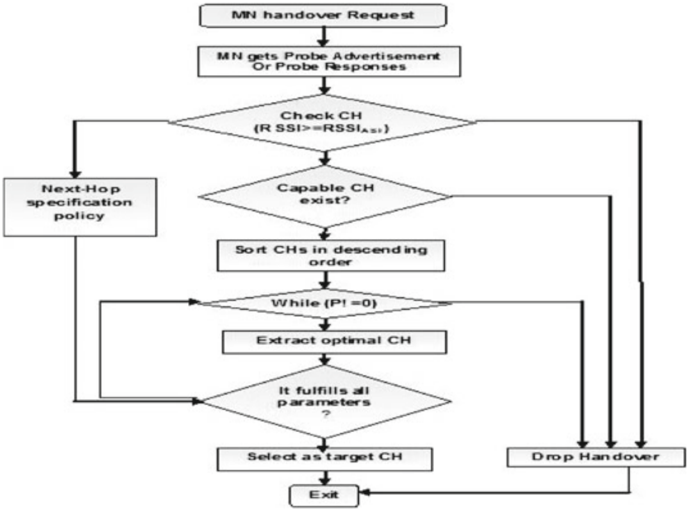

MN handover Request
MN gets Probe Advertisement Or Probe Responses
Check CH (R SSI>=RSSI_AST)
Next-Hop specification policy
Capable CH exist?
Sort CHs in descending order
While (P! =0)
Extract optimal CH
It fulfills all parameters ?
Select as target CH
Drop Handover
Exit

Fig. 5 Handover decision procedure

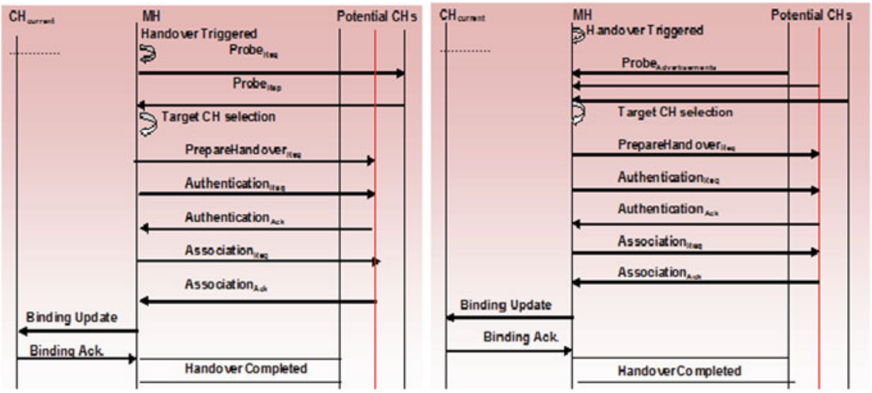


Fig. 6 Reactive and proactive handover procedure

4 Conclusion

MANET allows mobile nodes to spontaneously form a network and share their services but due to the network's dynamics and resource constraints of wireless nodes, It requires adequate support for service discovery. An effective cluster formation and power efficient cluster head selection mechanism based on reputation and mobility have been proposed in this work. Certain policy framework has been explored to achieve balanced clusters, adaptive service discovery and best cluster selection for a freely roaming mobile node. These policies have been enforced on all handover phases resulting in optimized handover with optimal resource consumption, reduced latency and interruption time (Fig. 6).

5 Future Scope

This mechanism is proposed to achieve required level of flexibility and adaptability to combat the network dynamics. Further exploration is required for a formal validation, based upon prototype implementation and performance evaluation. There is a scope for additional policies which can be explored and implemented as per requirements to increase the effectiveness of proposed scheme. An effective mechanism for location prediction can be incorporated in this framework to assure an optimized handover.

References

1. Berlin Hency V, Aravind Prasad S, Kannan YRA, Sridharan D (2011) A Novel approach for handover avoidance and Qos improvement in Wlan. Int J Distrib Paral Syst 2(5):77–86
2. Cynthia J, Sumathi V (2011) Adaptive service discovery protocol for mobile ad hoc networks. Eur J Sci Res 49(1):6–17

3. Faisal M, Nawaz MK (2012) An enhanced scheme for reducing vertical handover latency. Int J Adv Comput Sci Appl 3(1):100–105
4. Gupta A, Vatsa AK (2011) Policy based fast handover mechanism for MANET. World Comput Sci Inf Technol J 1(8):339–345
5. Jailani K, Osman G, Mohamed F, Suhaidi H (2011) Node selection based on energy consumption in MANET
6. Khan R, Vatsa AK (2011) Detection and control of DDOS attacks over reputation and score based MANET. J Emerg Trends Comput Inf Sci 2(11):646–655
7. Kumar P, Vatsa AK (2011) Novel security architecture and mechanism for identity based information retrieval system in MANET. Int J Mobil Adhoc Netw 1(3):68–72
8. Mishra S, Manish S (2011) Efficient secure clustering protocol for mobile ad-hoc network. J Glob Res Comput Sci 2(9):39–45

Trusted Control Information in the Application of Cold Chain Logistics

Li Li, Jia Yingqian, Zhao cuijian and Wu hao

Abstract Cold chain logistics as an research respect of Internet of Things, data mining technology is introduced into the cold chain logistics information processing. Using credible control classification algorithm, the information data was to expression, quantify, extraction and processing in the cold chain logistics system. The algorithm accuracy and the average usage time is to meet the demand by testing. The process of research and simulation analysis was to provide a simulation approach for the cold chain logistics in the Internet of Things applications and design, that has practical significance for speeding up applications the Internet of Things, especially to applications problems in cold chain logistics environment.

1 Introduction

With the arrival of the information age, the intelligent of the whole society ecosystem, which is the intelligent identification and management of all items, in order to achieve the increasingly strong demand for human society and the physical system integration, research and development of Internet of Things that has caused wide attention. In this context, the Internet of Things (Internet of Things, IoT) is a large virtual network through various access technologies that will be massive electronic devices and the Internet interconnection. The Internet of Things all items by radio frequency identification (RFID) technology, automatic identification, connect, and share of goods and information via the Internet, after the aggregation of heterogeneous information to achieve intelligent identification and management. Cold chain logistics, as an aspect

L. Li (✉) · J. Yingqian · Z. cuijian · W. hao
Shijiazhuang Institute, Shijiazhuang, hebei province of China
e-mail: lili2008mail@163.com

J. Yingqian
e-mail: cat_ll_tutu_2002@yahoo.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_65,

of the Internet of Things, which is another wave of the information industry after computers, the Internet and mobile communication network.

The cold chain logistics is a huge system and a systematic project to reduce food losses, and refers to meat, milk and other refrigerated food to ensure food quality, production, storage and transportation, sale, in the all aspects of the consumer provisions, always in the low-temperature environment. Before the practical application, the key technologies of the cold chain logistics system would be studying, which will reduce losses, improve efficiency, and improve food safety and transport quality. Thus contribute to the formation and development of cold chain logistics chain and the whole of Things.

System simulation can reveal the key technologies and key parameters of the cold chain logistics system in the Internet of Things. The credible control was used to expression, quantify and extraction information, to provide a simulation approach and simulation analysis for cold chain logistics in the design applications of the Internet of Things, for speeding up the development process of the Internet of Things, present, especially in the Internet of Things in the cold chain logistics environment in the existence of some problems with certain practical significance.

2 Information Intelligent Identification and Management

In recent years, the rapid development of the WSN(Wireless Sensor Networks, WSN) technology research and application in Internet of Things, mainly focused on the WSN in different application environments, some of the basic issues less involved in the WSN data communication.

2.1 Information Intelligent Identification and Management in Internet of Things

Follow-up study of WSN gradually extended to the field of semantic interoperability and information integration, semantic interoperability and information integration are an important feature of the Internet of Things, the authors believe that WSN is one of the basic technology of the Internet of Things. Therefore, the Internet of Things will certainly continue the research questions of the WSN. Therefore, before practical application, the key technologies research of WSN will reduce the difficulties of deployment in the enterprise and application of the cold chain logistics, and thus contribute to the formation and rapid development of the entire cold chain logistics industry chain and physical interconnection [1, 2].

In Internet of Things, object computing, storage and processing capabilities is different, types of objects from simple RFID RF card to have a strong ability video sensor. The information collected is different, how to express, storage, retrieval, sharing, and processing this information is a major problem facing the Internet of Things. WSN was applied only in some small-scale applications, in these applications, information processing can be simply divided into foreground and background processing. Internet of Things, with the increase of application sizes and application types, the information calculation processing and storage is an urgent need to address the problem in the Internet of Things, we trusted information exchange to transfer control technology applied to information computing and storage to solve the expression of information storage, retrieval, sharing and processing.

2.2 Information Intelligent Identification and Management of Cold Chain Logistics

Intelligent information processing in the cold chain logistics must be considered the computing power of object. Because of the limited computing power of sensor nodes can not be complex information processing to the sensor to complete. Intelligent information processing, computing tasks should be to the gateway with a strong computing power, or other network components to complete. Current intelligent information processing research focuses on the dynamic Bayesian networks, extended Kalman filter, DS (Dempster-Shafer Theory) evidence theory and rough set theory. Dynamic Bayesian network-based of intelligent information processing is evolved from the Bayesian networks, dynamic Bayesian network in the original Bayesian network added the new timeline. Dynamic Bayesian networks belong to a slightly more complex dynamic spatial model, which is similar but simpler to the hidden Markov chain and the Kalman filter model. These are do not consider the characteristics of energy restriction of the Internet of Things, the computational overhead of the system are not well decreases.

In scale application, the wireless sensor networks are some small-scale applications, mainly due to can not be information exchange between systems, or information exchange will pay a terrible price, the power consumption of the information transmission. The Internet of Things made precisely in order to solve the problem of the WSN scale and information processing, cold chain logistics as a research on the Internet of Things, large-scale problems and information processing is the key issue to limit the development of cold chain logistics.

3 Information Generalization and Summary Description

3.1 Information Access

In the cold chain logistics system based on the Internet of Things was hidden in the database for decision-making of the commercial, research and other activities to provide the required knowledge. Information for classification mining, the establishment a descriptive model of the known data categories or concepts, to obtain the information in the data collection is the basis of information processing.

Before classification, the first ready for data mining. Generally require the following pretreatment data: data cleaning, analysis, data conversion, which to help improve the classification accuracy, efficiency and scalability.

Noise and abnormal data on the data set, data cleansing to help remove noise in the data, effectively reducing the learning process that may appear contradictory situation. Since the many properties of the data set may be unrelated the data mining tasks [3].

For example, a cold chain logistics database, cf. Table 1.

For example: record cargo transport application form to fill out the number of weeks (attributes), it is possible with the application successful without describe; In addition, some properties may also be redundant. Therefore, by data analysis, help to eliminate irrelevant or redundant attributes in the learning stage. For some generalization to a higher level value conversion. : for instance, attribute "humidity" values can be generalized in a number of discrete interval, in addition, such as: low, medium and high.

3.2 Information Representation

In this paper, the attribute value to the information expression, such a theory approach will help to effectively reduce the number of object classification [4]. Using information gain method to help determine the appropriate attributes used in each node generated [5, 6]. Can choose the properties with the highest information gain as the test attribute of the current node, use the property for the current sample set partitioning, will make every book produced by concentration of the different categories mixed reduced to a minimum.

To calculate the information gain of each attribute, and pick out the information gain the greatest attribute as a given combination of test attributes, the resulting node is marked as the property, the corresponding branch bring by the different values of this property, each branch represents a division sample subset.

Table 1 The cold chain logistics database

Rid	Temperature	Humidity	Communicate	Power consumption	Status
1	o°	1°	yes	fair	Normal
2	5°	5°	yes	fair	Alarm
…	…	…	…	…	…
100	20°	22°	no	fair	Normal

Set D be a set of d data samples. Another class attribute can take m different values corresponding to the m different categories Vi$^{i\in\{1,2,3,\ldots,m\}}$。. Assume that di is the number of samples in the category Vi, then the amount of information a given data object classification as follows:

$$I(d1, d2, \ldots, dm) = -\sum_{i=1}^{m} pi \log(pi) \tag{1}$$

where pi is the probability belongs to the category Vi of any one data object; based on di/D calculation. The log function is base 2.

Set an attribute A take n different values $\{a1, a2, \ldots, an\}$. Based on attribute A, collection D is divided into n subsets $\{D1, D2, \ldots, Dn\}$, where Dj include take aj value sample of attribute A in D data collection. Attribute A is chosen as the test attribute, set dij is belongs to Vi category the number of samples in subset Dj. Then use the information entropy of attribute A division of the current sample set is calculated as follows:

$$E(A) = \sum_{j=1}^{n} \frac{d1j + d2j + \cdots + dmj}{d} I(d1j, d2j, \ldots, dmj) \tag{2}$$

which $(d1j + \cdots + dmj)/d$ is treated as the j subset weights, it is the total number of samples in the collection of attribute A take aj value of all subset number of samples divided by D. E (A) evaluates the smaller the subset division result is the better. For a stator set Dj, its information entropy

$$I(d1j, d2j, \ldots, dmj) = -\sum_{i=1}^{m} pij \log(pij) \tag{3}$$

which $pij = dij \Big/ |dj|$. is a subset of the probability of dj any one data sample belongs to the class Vi. Based on attribute A, the information gain obtained by the corresponding sample collection divided as follows:

$$Gain(A) = I(d1, d2, \ldots, dm) - E(A) \tag{4}$$

Gain (A) is the entropy of information reduction value obtained in the sample attribute A. For each attribute to calculate the information gain to choose the information gain the greatest attribute as test property of a given set of D give rise to a branch node, the node is marked for the corresponding property, and according to the different values of the property divided into subsets.

For non-class attributes, in the information gain exceeds the category attributes, information gain is not as the sole judgment conditions, the need to recalculate the weights of the property based on after the probability of taking the maximum principle, the right to adjust the value to update the information gain. Information classification is not determined by the information gain of the class attribute, but determined by the attribute weights.

Definition 1 Let Properties An as non-class attribute, $\{p(c1, c2, \ldots, cn)\}$ as probability of occurrence, the $\{c1, c2, \ldots, cn\}$ is the n different classes to be determined, then the weights of q is:

$$q = p(x|ci)p(ci) \Big/ p(x) \tag{5}$$

3.3 Information to Quantify and Extract

Cut into several categories of data, using the concept of equivalence class, carved out of the equivalence class is a group classification; further analysis the characteristics of each classification, can get a classification of characteristic rules. This analysis has practical significance, for example: analysis of a large number of alarm data can be drawn from a variety of alarm data and reaction characteristics.

Description of the rules of classification characteristics of a four-tuple lw = <T,S,C,G>, where T is a finite set of a group of objects (or instances), called on the domain, with n objects, T can be expressed as follows: T = {t1, t2, . . ., tn}; S is that the combination of attributes, with m attributes, then S can be expressed as: S = {s1, s2, . . ., sm}; range of the property set C can be expressed as: C = {c1, c2, ..., cm}, range c1 = {ci1, ci2, ..., cik}; G is a function of t and s, cij = f (ti, si).

Assume a collection has 14 elements in T, ie, 14 records in the table; five elements in the set S, the field names of those in Table 5, s1 = 'the rid'; c21 = f (t2, s1) ='<00'

Definition 2 Set D be a database, P is the total number of records, X is $S'(S' \subseteq S)$ the equivalence class, Sx is the record number of X, then P′ = Sx, /P is the classification support of the equivalence class X.

For the above example, do the communication-based division, can be divided into: communication = no, that is E1 = {r1 and r2, r3, r4, r8, r12, r14}; communications = yes, that is E2 = {r5, r6, r7,r9, r10, r11, r13} two classifications, classification support for 7 and 7.

Definition 3 Set X be an equivalence class of S, S is the complement of S on St,$\{\overline{S}\} \in St$. B is a subset of S, Y is the equivalence class based on the B, St is the number of records in B, then Q = St/Sx is the characteristic confidence of equivalence class X.

Our concern is the classification to support a greater degree (greater than a certain threshold). This classification feature is defined as the classification of characteristic rules.

The rules of the classification feature can be described as $B \to Y(Sx|St)$, where B is the category; Y is characterized; Sx is classification support; St. is characterized confidence. Cases of communication category, not the characteristics of communication confidence level of 7/7, the rules of classification feature can be expressed for the communications →no Communications (7,7).

4 Information Processing Based on the Credibility of Control

4.1 The Credible Control Classification Algorithm

Algorithms: the attribute-based credibility control classification algorithm, based on user data mining request, qualitative description of the mining in relational databases.

Input: 1.the relationship database DB; 2. data mining command DMQuery; 3. a set of properties alist; 4. attribute ai generalization, attribute generalization threshold calculation for each attribute ai athresh (aiThe); 5. the property right value to adjust; 6. data classification operation.

Output: Primerelation contains a qualitative concept description based on the alist attribute set.

The algorithm is described as follows:

```
get_data(DMQuery,DB,W_relation);
scan w_relation to count t_values(ai);
 for each ai in alist where t_value(ai)>athresh(ai)
                  if(Gain(ai) is no)or(aithat is A higher level concepts are expressed by other attributes)
                  removeattribute(ai,alist);
for each ai in alist where t_values(ai)>athresh(ai)
            while(t_values(ai)>athresh(ai))
            generalize(ai, Gain(ai), t_values(ai), W_relation)
          for each ai_qweigh>yvalues
      qweigh recomputer and qweighx=qweigh;
      else qweigh;
working ai_class;
```

4.2 Algorithm Testing

The data is in the cold chain logistics database, after learning the training samples, by test samples for testing. Collection of test samples as follows, such as a cold-chain logistics database, cf. Table 2.

The information needed to classify the given test sample humidity as follows:

$$I(S1, S2) = 0.940 \tag{6}$$

Table 2 Test cases library

rid	Temperature	Humidity	Communicate	Power consumption	Status
1	<0°	high	no	fair	no
2	<0°	high	no	high	no
3	0–15°	high	no	fair	yes
4	>15°	medium	no	fair	yes
5	>15°	low	yes	fair	yes
6	>15°	low	yes	low	no
...	...	...	...	...	...
995	0°–15°	high	yes	fair	yes
996	>15°	medium	no	high	no

Calculate the information entropy of the property, when the humidity is high,

$$I(S11, S21) = 1.000 \tag{7}$$

when the humidity is medium,

$$I(S12, S22) = 0.918 \tag{8}$$

when the humidity is low,

$$I(S13, S23) = 0.811 \tag{9}$$

$$\begin{aligned} E(T) = \frac{S11 + S21}{S} I(S11 + S21) + \frac{S12 + S22}{S} I(S12, S22) \\ + \frac{S13 + S23}{S} I(S13, S23) = 0.911 \end{aligned} \tag{10}$$

Attribute information gain as

$$Gain(S) = I(S1, S2) - E(S) = 0.029 \tag{11}$$

Similarly, the temperature gain is 0.245; communication information gain is 0.3425; the power of information gain is 0.048. Here the communicate information gain is greatest, to classify the data firstly according to the communication information gain, after the classification according to temperature information gain.

When the need to further distinguish the abnormal data is appear, when the temperature is higher than 150 can not be further classified according to the power consumption, based on the after maximum probability principle of equivalence classes of re-adjust the weights of the attributes, according to the power consumption of the fair that the weight is 0.033, according to the power consumption is high, the right weight of 0.011, significant compared to take the right weight is greatest as a classification basis (Table 3).

Table 3 The Algorithm average time

Data set n	Average time (ms)	Improved the average time (ms)	Weighted average time (ms)
314	413.1	355.1	356.9
598	419.6	361.2	362.5
996	439.7	398.3	400.1

4.3 Performance Analysis

We apply the above algorithm, data mining database of cold-chain logistics system, the accuracy of the algorithm in more than 80.1 %, to reach the needs of the system. The amount of data for the data set comparison algorithm time, as follows

5 Conclusion

In information processing, in-depth study of information acquisition, expression, quantify, extraction, and the process of reasoning. Credible information exchange technology simulation study in the study on the basis of existing information processing method, combined with the characteristics of the cold chain logistics system uses a combination of data mining and information filtering. System simulation can reveal the key technologies and key parameters of the cold chain logistics system in the Internet of Things, the findings provide a reference for the actual cold chain logistics system design and application.

Acknowledgments This paper is sponsored by the Shijiazhuang College, Dr. Fund and the scientific technological research and development fund of Shijiazhuang City under the Grants No. 11BJ005 and No. 11113981

References

1. Zhu B, Longxiang Y (2010) Thought and strategy of things. Commun Univ 31(11):2–9
2. Baoyun W (2009) Of things review. Electron Meas Instrum 23(12): 1–7
3. Wenbin X, Zhang K-L (2000) Attribute classification-based data mining method. Small microcomputer system. 21(3):305–308
4. Hu B (2006) Attribute reduction and rule generation algorithm of data mining. Huazhong University of Science and Technology, China
5. Zhangliang C (2009) Decision-making based on data mining prediction model application and research. Manag inf 12(1): 57–59
6. Wei J (2009) Classification algorithm for data mining analysis. Comput Knowl Technol 5(1):18–19

PHY Layer Considerations for Real Time Multimedia Signal Transmission in Wireless Medium

S. M. Koli, R. G. Purandare, S. P. Kshirsagar and V. V. Gohokar

Abstract For real time multimedia signal transmission in wireless channel, the effect of packets lost due to error and due to congestion, end-to-end delay during the transmission which in-turn will affect the parameters of wireless channel like reliability, quality of service (QoS); transmission rate, transmission delay etc. are required to be considered. Most of the real-time applications uses user datagram protocol (UDP) as their transport protocol. For better end-to-end performance of wireless network, improvements are required in the existing protocol of physical (PHY) layer. This paper surveys the literature on real time communication over wireless and suggests adaptive error control mechanism for reliable transmission. The suggested adaptive forward error control (AFEC) method results in achieving the desired recovery rate at the receiver without using any retransmission mechanism. The network simulator NS-2 is used to evaluate the same. The simulation results indicate that the suggested method improves end-to-end performance by increased packet delivery ratio (PDR) by reducing end to end delay for real time multimedia signal in the wireless network.

S. M. Koli (✉)
Electronics and Telecommunications Engineering Department, Smt. Kashibai Navale College of Engineering, Vadagon (Bk), Pune 411 041, Maharashtra, India
e-mail: smkoli.skncoe@sinhgad.com

R. G. Purandare
Electronics and Telecommunications Engineering Department, Vishwakarma Institute of Information Technology, Kondhwa (Bk), Pune 411 048, Maharashtra, India
e-mail: purandare.radhika@viit.ac.in

S. P. Kshirsagar
AnchorTek Techno Consultancy Pvt. Ltd., Pune 411 048, Maharashtra, India
e-mail: shirish@anchorteksys.com

V. V. Gohokar
Electronics and Telecommunications Engineering Department, Shree Sant Gajanan Maharaj College of Engineering, Shegaon, Maharashtra, India
e-mail: vvgohokar@ssgmce.ac.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_66,

Keywords Real time multimedia signal · End-to-end delay · PDR · Network throughput · AFEC · NS-2

1 Introduction

The rich integrated multimedia services the smart handheld devices of user and demands the wireless link to be robust to error [1]. For the real time multimedia signal transmission in wireless medium, available radio resources are limited and hence there is a need of development of adaptable radio interfaces to adapt the variations in the wireless link, in order to optimize network performance [2, 3] which will result in good end-to-end wireless network performance.

The end-to-end flow and congestion control mechanisms are required to analyze the end-to-end performance of a protocol that has been designed and optimized for wired and wireless networks [3–5]. The deleterious effects of noise, interference, jamming, fading, and other channel impairments [1, 6] can be effectively reduced by using channel coding and interleaving techniques [7]. Section 2 gives a brief overview on channel coding. Various wireless transmission techniques related to the Internet transport layer protocols available in the literature are evaluated in Sect. 3. Section 4 describes implementation of different modules proposed. The article highlights the results with discussions in Sect. 5 and ends with conclusions in Sect. 6.

2 Channel Coding

The idea of channel coding is to introduce some controlled redundancy into the original data to allow reconstruction of damaged blocks at the receiver. The redundant data is generated from the original data using techniques from coding theory, where two such techniques are well known in the literature as feed-forward error correction (FEC) and automatic repeat request (ARQ) [6, 8–10].

(a) FEC

The channel encoder in the transmitter employing FEC, accepts block of ‘k’ message bits or packets and adds *redundancy* R according to the prescribed rule as $\mathrm{R} = (\mathrm{n}-\mathrm{k})$ in the controlled manner. The redundant data can be used to recover lost source data at the receivers. A receiver can reconstruct the original source data once it receives a sufficient number of packets. The combined goal of the channel encoder and decoder is to minimize the effect of channel noise. No communication is done by the receiver with transmitter after decoding, hence FEC require simplex link. FEC schemes are further classified as media independent FEC and congestion control [2, 6, 11, 12].

(b) ARQ [8, 9]

Upon the detection of an error in the transmitted packet, the receiver requests the sender to repeat transmission of the lost packets (LP), which necessitates the use of the return channel and results in a great network load and hence are typically not acceptable for live audio and video applications and out of the scope of this paper. In Sect. 3, various Internet transport layer protocols are revisited.

3 Internet Transport Layer Protocols

The transport layer [8, 9, 13] provides end-to-end segment transportation of messages and are reassembled back into the original message at the destination nodes. Examples of transport protocols are transmission control protocol (TCP), the user datagram protocol (UDP). TCP/IP and UDP/IP are the core of today's Internet transport layer [3, 9, 13, 14].

(a) TCP

TCP is the reliable transport layer protocol and most Internet applications today rely on the TCP. The TCP/IP is the most popular protocol suite where stream data transfer reliability, efficient flow control, full-duplex operation and multiplexing is required, but it is not suitable for real time transmission.

(b) UDP

UDP is connectionless transport layer protocol which doesn't require connection establishment prior to data transfer and runs on top of IP networks. UDP data units are called data-grams; also referred to as blocks. The applications where reliability is not critical or with strict transmission delays constraints such as real-time traffic are carried by the UDP/IP protocol. Here the datagram is send with the hope that receiver will be able to handle it. Hence, the UDP protocol is an unreliable protocol and it is suitable for broadcast of data [6, 8, 10, 15]. But UDP has a disadvantage of loosing many packets. But it is not critical for real time applications. It does not provide sequence number management while exchanging data-grams and does not guarantee orderly transmission. It also does not offer capabilities for congestion or flow control [13]. The protocol in charge of providing sequence number control is the RTP protocol running on top of UDP (RTP-on-UDP). UDP contains no ACK mechanism; therefore, the lost data-grams can be recovered only by lower or upper layers, including the application layer [14]. Each UDP datagram is composed of a header and a payload (user data). In the payload the data coming from the layer above is encapsulated. UDP takes messages from the application process, then adds the source and destination port number fields, the length field and finally the checksum. The resulting segment is passed to MAC layer where it is encapsulated into an IP datagram.

When the datagram is received at the receiver it is divided into an IP header and IP payload. The latter is passed to the transport layer i.e. the UDP layer. UDP then uses the port number contained in its header to deliver the data to the correct application. Therefore, if a datagram does not arrive to the receiver, there is no possibility to recover it or to ask for a retransmission, because the receiver does not have any information about the sent packets. It has only the source port which takes into account the equivalent field at the IP header to protect the header.

RTP-on-UDP [14]

RTP is a generic transport protocol which is independent of applications. But is implemented at the application layer and is designed to handle real-time traffic on the Internet. RTP does not have a delivery mechanism; it must be used with UDP. RTP stands between UDP and the application program. The main contributions of RTP are time-stamping, sequencing and mixing facilities. The RTP payload is filled with data coming from the application layer. Then this RTP packet plus the RTP header are passed to the lower layer, the UDP layer. The UDP layer takes this RTP packet and adds the UDP header. Then this UDP datagram is passed to the IP layer who adds also its header. Finally the IP datagram is passed to the physical (PHY) layer which also adds its header. Applications typically run on RTP-on-UDP to make use of its multiplexing and checksum services. Therefore, RTP protocol provides end-to-end delivery services for data with real-time characteristics, such as audio and video [4]. But, note that RTP itself does not provide any mechanism to ensure timely delivery or provide other quality-of-service guarantees, consequently RTP have to rely on lower-layer services to do it. A RTP packet consists of a fixed RTP header and a RTP payload.

The receiver has, thanks to RTP, enough information to manage the packets. But suppose that a packet have been lost somewhere in the network due to congestion or to link, then once the sender notice that a packet has been lost, it has to handle it somehow. Some intermediate solution providing more reliability to the communication between applications running on top of UDP and at the same time support its fast connection quality using already existing protocols is required. FEC coding commonly serves the purpose and included in the PHY layer design of wireless links [15]. This error control mechanism attempts to minimize the visual impact of LPs at the destinations [1, 3, 13].

FEC can be applied both at the bit-level and on the packet-level. Normally packet-level FEC is applied to end-to-end communication whereas bit-level FEC is used on a specific link. Packet-level FEC [3, 7, 8, 10, 16, 17] is chosen in this paper because more information about the requirements of the application can be used. Suppose that an amount of k data packets has to be sent in a block. To this data, $R = (n - k)$ (RPs) or parity check packets are appended. The RPs are calculated using block coding theory. Therefore, finally a total of $n = k + R$ packets are sent per block.

Recovering 'k' packets at the receiver is achieved if no more than 'R' packets are dropped. If more RPs are added than needed, then the network load will be unnecessarily increased. Nevertheless, if less RPs than needed are added then the LPs will not be recovered and also the network will be loaded with redundancy traffic. Therefore, there is a clear compromise between the obtained throughput and the network load. The relation between the packet delay and throughput with the network load can be understood with the help of Fig. 1 [9].

From Fig. 1a it is clear that when the load is much less than the capacity of network, the delay is at a minimum, whereas when the load reaches the network capacity the delay increases sharply. The delay becomes infinity when the load is greater than the capacity. Delay has a negative effect on the load and consequently the congestion. The wireless network throughput is the average rate of successful bytes delivery over a channel.

Figure 1b shows that the throughput almost linearly increases if the network load doesn't exceed the network capacity resulting into no congestion. But the throughput decreases logarithmically reaching almost to zero if the network load exceeds the network capacity resulting into congestion. From the point of view of the packet loss probability (PLP), the added redundancy makes it to decrease. The amount of RPs has to be chosen trying to get as small packet loss rate (PLR) as possible. Normally RS codes are used, but another such as parity or Hamming codes could be also used [18]. FEC is independent of the nature of the application data, but is implemented on the application layer. The FEC packet is obtained by placing the FEC header and the FEC payload in the RTP payload [9], as is shown in Fig. 2. The FEC payload is composed by the application data and the FEC header is constructed by placing on it the redundant packets and is supposed to be sent to the lower layer at the UDP/IP stack, the RTP protocol.

All the concepts considered so for can be represented by block schematic in relation with the TCP/IP network model of computer communication network as shown in Fig. 3.

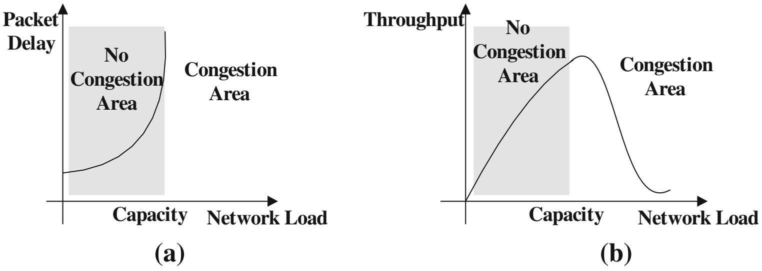


Fig. 1 Relation between the packet delay and throughput with the network load

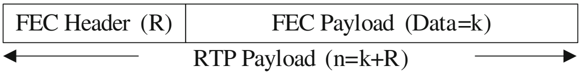


Fig. 2 Encapsulation of FEC-on-RTP protocol

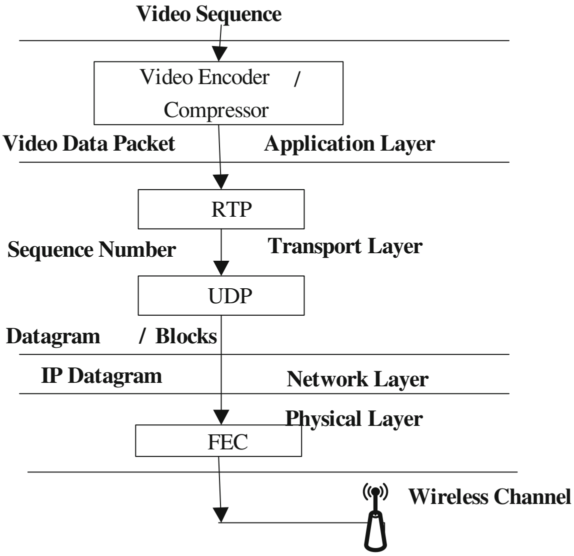


Fig. 3 Video over wireless channel in relation with network model

(c) Adaptive Forward Error Correction (AFEC)

Good end-to-end wireless network performance will not be possible without a truly optimized, integrated and adaptive network design [1, 3]. Hence, instead of fixing a level of overhead that can cope with worst-case conditions, adaptive error control mechanisms let the error protection vary as the conditions vary i.e. the overhead is always adapted to the current conditions of channel. As has been mentioned, the UDP receiver does not have any information about the sent packets. RTP running on top of UDP provides sequence number value to each sent packet. Once the receiver has the sequence numbers it can know whether a packet has been lost or not. But, if there is no retransmission mechanism in UDP, how could this LPs be recovered?

In particular, in this paper, the number of RPs will be varied in response to the changing PLP trying to keep a specific PLP quality after decoding. Thus, an adaptive feature is added to FEC and then it is renamed as AFEC. The idea of AFEC is to inject an amount of redundant packets in every sent block (or datagram) in order to achieve a desired recovery rate at the receiver without using any retransmission (of data) mechanism. The specific amount of redundancy is updated by the sender based on channel loss probability measurements as shown in Fig. 4, which explains the

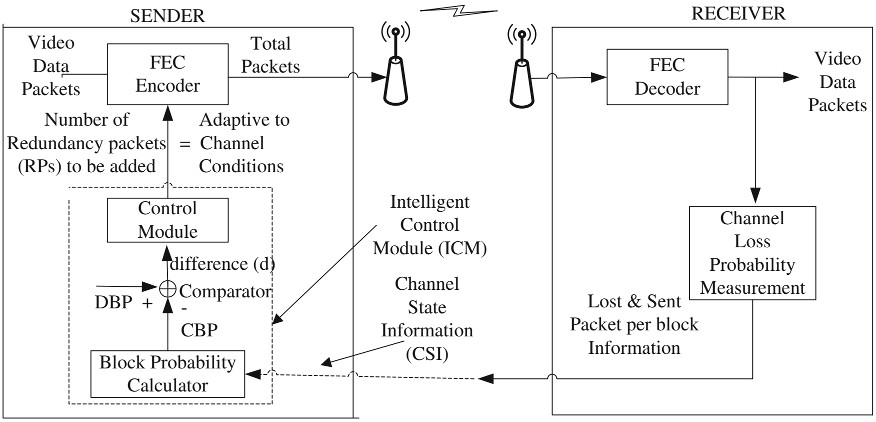

SENDER
Video Data Packets
FEC Encoder
Total Packets
Number of Redundancy packets (RPs) to be added
=
Adaptive to Channel Conditions
Control Module
difference (d)
Comparator
DBP
+
-
CBP
Block Probability Calculator
Intelligent Control Module (ICM)
Channel State Information (CSI)
RECEIVER
FEC Decoder
Video Data Packets
Channel Loss Probability Measurement
Lost & Sent Packet per block Information

Fig. 4 Concept of adaptive FEC and its implementation

method to implement proposed AFEC. The block probability calculator along with control modules is called as intelligent control module (ICM). FEC along-with this ICM is called as AFEC. This complete wireless system is simulated using NS-2 [19].

In the simulation, the behavior of ICM is checked for the number of LPs per block is constant, for adding loss probability in a link, further increasing loss probability to understand the system behavior and adding competitive sources to cause congestion. All these four simulation cases are tested for the three different ICMs proposed in next section. The information regarding number of lost packets due to error is referred as channel loss probability measurement and indicates lost as well as send packets per block information and is also called as channel state information (CSI). These measurements are made by the receiver when a new block is received. The receiver sends this information as CSI to the sender who is the one in charge of updating the necessary amount of redundancy. Thus, if a block turns out to be corrupt by the loss of some of its packets, the receiver will be capable of recovering the lost information, if no more than the amount of introduced RPs has been lost or corrupted [3, 11]. Thus, the RTP has been chosen as upper level protocol to provide flow control in AFEC technique to provide error management, which answers the question raised above. To obtain the appropriate number of RPs from the calculated block probability (CBP) various ICMs or feedback control systems can be used [17, 12]. In this paper ICMs such as a Delta intelligent control module (DICM), a Adaptive Delta intelligent control module (ADICM) and a Adaptive Delta Sigma intelligent control module (ADSCM) are proposed, details of which is explained in Sect. 4.

Instead of AFEC/UDP other solutions have been proposed in the literature as reliable UDP (RUDP) or UDP Lite [20, 21] but they are out of the scope of this paper. RUDP [20] is suitable for transport telecommunication signaling and is layered on the UDP/IP. UDP Lite was proposed to prevent packets loss at the receiver side if channel errors are only located on the packet payload. Therefore, errors detected in the packet header result in a discarded block whereas errors detected in the packet payload does not result in a discarded block [21]. Enhanced AFEC (EnAFEC) [22] is also proposed but it estimates the suitable smoothing factor value to determine the average queue length according to the packet loss rate over the burst wireless error network.

4 Implementation of ICM

The CSI from the receiver is used by the sender to calculate the optimal number of RPs to be added to FEC code. Feedback information is updated only when a block has been totally received. The sub-block of sender in Fig. 4, called ICM does following calculations.

- CBP calculation. It is calculated by the block probability calculator using Eq. 1 [23].

$$P_{(\mathrm{CBP})}(n,k)=\sum_{i=0}^{k-1}\binom{n}{i}(1-\mathrm{P})^{\mathrm{i}}(P)^{n-i}\left(\frac{\mathrm{n}-i}{\mathrm{n}}\right) \tag{1}$$

  where: ‘P’ is the estimated packet loss probability without FEC, ‘n’ is the total number of packets per block, ‘k’ is the data packets per block and ‘$\frac{n-i}{n}$’ which gives the average number of lost packets when the losses cannot be recovered. Therefore, with this factor the Eq. 1 gives the average packet loss probability and without this factor the formula gives the block loss probability.
- CSI calculation is done with the help of LP per block and sent packet per block information estimated by the channel loss probability measurement unit at the receiver. With respect to Eq. 1, CSI can be mathematically represented as, $\hat{P}_{CSI}(n,k)$.
- DBP calculation. It is the desired block probability and is set to 0.005 because it is desired to achieve a small block loss probability in the simulations. With respect to Eq. 1, DBP can be mathematically represented as, $\hat{P}_{DBP}(n,k)$.
- Error d = DBP − CBP and is calculated by the comparator. This error could be positive or negative based on CBP and is represented by $\pm\delta$ (Delta). This error can be mathematically represented as, $d(n,k)$.
- RPs calculation. The error ‘d’ is applied to the control module which calculates number of RPs to be added by FEC encoder and acts as adaptive parameter for AFEC. These RPs (R) are inserted into the data block ‘k’ resulting in a total of ‘n’ packets to be sent to the network.

(a) Delta Intelligent Control Module (DICM)

The DICM is the simplest control module and uses only the information about the number of LPs and RPs per block. The operation of this system is based on a step

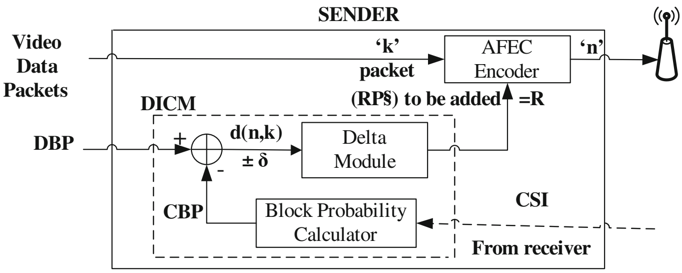


Fig. 5 Delta intelligent control module (DICM)

by step increments or decrements of the ICM output i.e. number of header packets (HPs), which depends on the error 'd' which is either $+\delta$ or $-\delta$. The number of header RPs in-turn will be incremented or decremented by one. Hence the name to the scheme is given as Delta module and is shown in Fig. 5.

dummy

Suppose that the number of LPs in the received block is denoted by the 'LP' variable and the number of HPs for the same block is denoted by 'HP'. Then working of DICM can be understood with the help of the flowchart as shown in Fig. 6. For the case LP > HP, the header is increased by one packet because it is quite probable that in the next block a similar amount of packets will be lost as uniform distribution

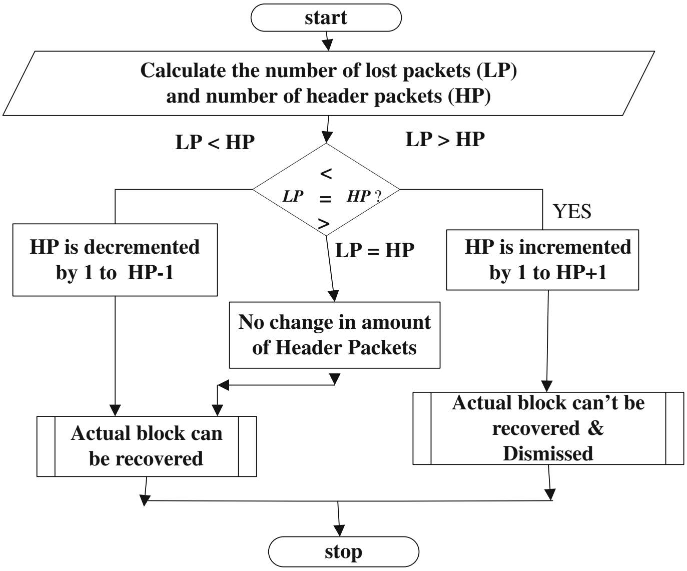

start
Calculate the number of lost packets (LP) and number of header packets (HP)
LP < HP
LP > HP
LP < = > HP ?
YES
HP is decremented by 1 to HP-1
LP = HP
HP is incremented by 1 to HP+1
No change in amount of Header Packets
Actual block can be recovered
Actual block can't be recovered & Dismissed
stop

Fig. 6 Flowchart showing working of DICM

of the network errors is used. As increment or decrement of the number of HPs is made one by one based on $+\delta$ or $-\delta$, this method is not the most optimal method.

(b) Adaptive Delta Intelligent Control Module (ADICM)

To make DICM optimal, increment or decrement in the number of RPs should be done in steps of size bigger than one packet in order to follow better the changes in the number of LPs per block. Additional HP size controller is added to the ICM of DICM which will increases or decreases the number of HPs by more than one, based on the error ‘d’ (i.e. $+\delta$ or $-\delta$) which in-turn controls the increment or decrement of HPs by more than one. This information is used to get the number of RPs to be added. Hence the technique is named as Adaptive DICM (ADICM). The complete schematic is as shown in Fig. 7.

The RPs (R) are introduced by the sender in every block in order to achieve the small number of dismissed packets in the receiver side after decoding as possible. In the proposed ADICM the relation between the output and the input of the system is given by a constant, called ‘γ’ and the system has to be tuned by fixing this gain γ based on the results of the simulations. The results of several simulations using NS-2 for the two main cases mentioned before are:

- For the first case, where the number of LPs per block is constant, the value of γ is increased starting from 0 until the header value exceeds the constant number of LPs per block and remains stable. The obtained value is $\gamma = 170$.
- For the case of variable number of LPs in a block the scenario chosen to tune the ICM is a lossy link with uniform probability distribution and loss probability equal to 0.03.

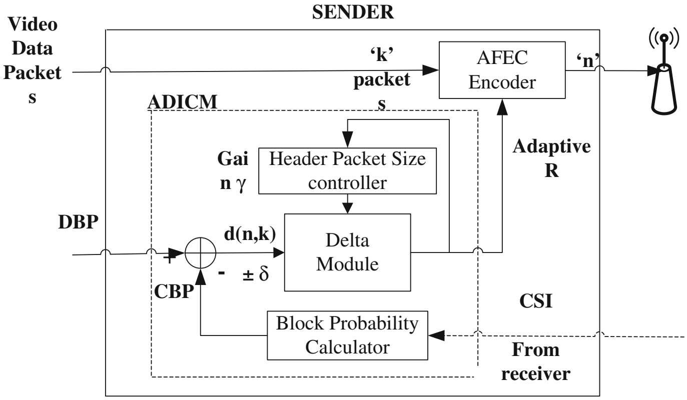


Fig. 7 Adaptive Delta intelligent control module (ADICM)

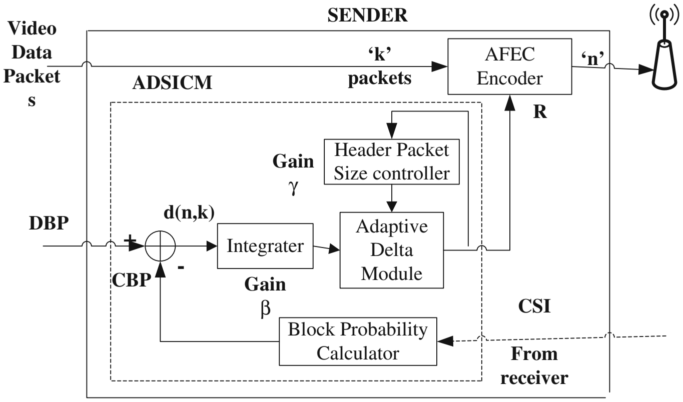


Fig. 8 Adaptive Delta Sigma intelligent control module (ADSICM)

(c) Adaptive Delta Sigma Intelligent Control Module (ADSICM)

The ADSICM introduces a integrator (a low band filter) in the ADICM system in order to hide the effect of the residual noise and follow better the changes in the number of LPs per block. The block diagram of such a new system is shown in Fig. 8.

In this case, both the integral gain β and the adaptive gain γ, has to be fixed. Also for this ADSICM system the choice of the parameters γ and β is made based on the results of the simulations trying to achieve as few as possible number of dismissed blocks and at the same time trying not to waste so much RPs. The use of an excessive number of RPs will the effect of load the network unnecessarily. Here also two cases are considered when the controller is tuned.

- For the first situation, where the number of LPs per block is constant, the values γ and β are increased starting from 0 until the header value exceeds the constant number of LPs per block and remains stable. The obtained values are $\gamma = 170$ and $\beta = 89$.
- For the case of variable number of LPs in a block the scenario chosen to tune the ADSICM module is the same that was chosen for the ADICM module, a lossy link with uniform probability distribution and loss probability equal to 0.03.

5 Results and Discussions

The AFEC system is implemented and simulated using Network Simulator 2 (NS-2) [19]. The complete experimental setup used for simulation is shown in Fig. 9. To generate this situation in the network one can set buffer size normally much higher (@300 packets), So that the congestion would not occur. For simulation of this

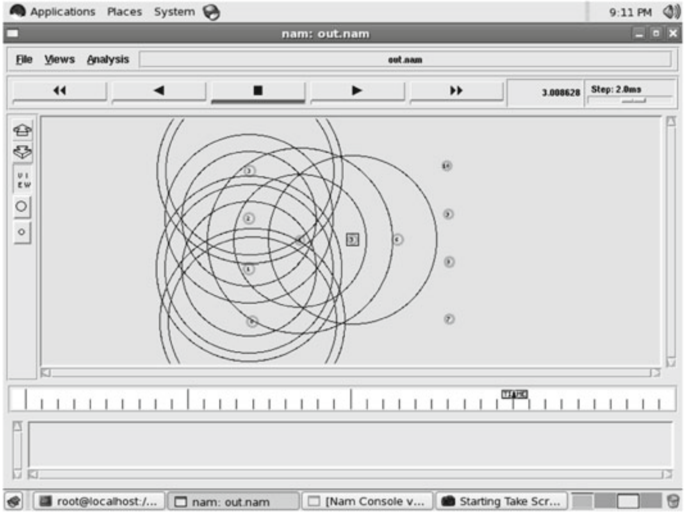


Fig. 9 Nam output: wireless video transmission system scenario

condition, different alternatives like keeping the number of LPs per block constant, introducing an uniformly distributed loss probability in the lossy wireless link and increasing loss probability in the critical wireless link (link between node 4–5–6) are considered. The situation of congestion can be created by setting the buffer size of intermediate wireless nodes too small (@100 packets), so that packets are lost.

Analysis is carried out by number of blocks sent and number of packets sent. The performance of the three ICMs of AFEC sender as described in Sect. 3, is evaluated using NS-2 and various results are plotted as shown in Fig. 10a–h. All the simulations for the three ICMs have been running separately and afterwards all the results have been summarized and plotted together. The system parameters required for simulation are set as shown in Table 1.

Now the scenario of congestion in nodes of wireless link is considered. The DBP is set to 0.005 for all the three suggested module of ICMs in AFEC. Figure 10a shows the EBP which is calculated from CSI and added to the DBP to generate error 'd'. This error 'd' is used by the FEC encoder to calculate the number of RPs based on the number of blocks 'k' of video data. The AFEC algorithm has to rectify this error

Table 1 Simulation parameters

Parameter	Value	Parameter	Value
Packet size	200 bytes	Gain γ for ADICM	170
Packet interval	0.004 s	Gain γ for ADSICM	170
Loss rate for error model	0.1 packets/s	Gain β for ADSICM	89
Number of data blocks k	16		

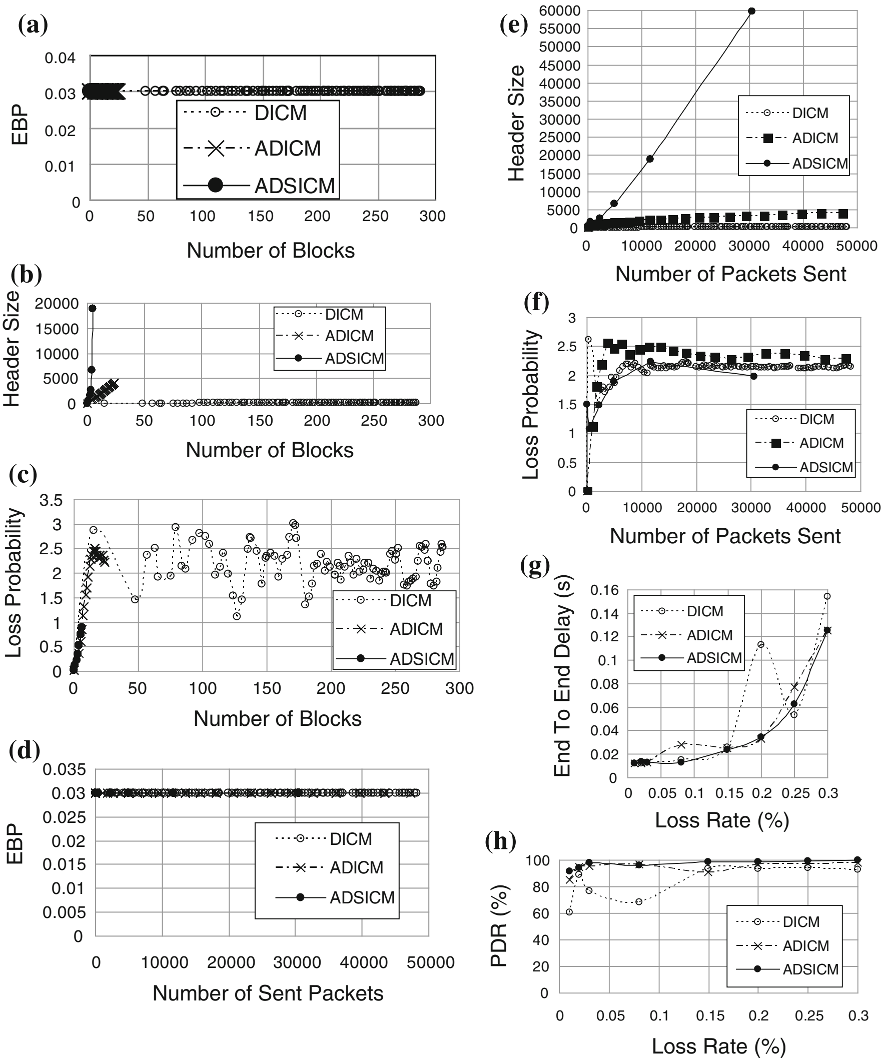


Fig. 10 **a** Number of blocks versus header. **b** Number of blocks versus header. **c** Number of blocks versus loss probability. **d** Number of sent packets versus estimated probability. **e** Number of sent packets versus header. **f** Number of sent packets versus loss probability. **g** End to end delay. **h** Packet delivery ratio (PDR)

after receiving the CSI form receiver by adapting the RPs accordingly. The simulation result shows that for DICM this error remains for all the blocks of data, for ADICM the error present only upto 25 blocks of video data whereas ADSICM rectifies the error within 10 blocks of data when transmission starts. Hence it is efficient. Based on the EBP, the ICMs computes the RPs header size. These RPs are added to the data blocks to generate ‘n’ packets to be sent. Figure 10b shows the size of HPs required by all the three ICMs.

Large header size is required by the ADSICM but during the starting phase only which is used to find the error and rectify it. Introduction of integrator in ADSICM reduces EBP further to negligible value, because ADSICM reacts to the increase in the block loss probability by increasing the number of HPs as shown in result. The two other ICMs needs HP to be changed for more number of blocks of data still after the communication starts. Figure 10c shows the relation between number of blocks of video data and loss probability. The packets are lost due to congestion in wireless channel. It is more clear that for ADSICM the packet loss probability is very less. Within 10 blocks of transmission itself it exactly computes the loss probability and receiver receives the data without error. Now the scenario of lossy wireless link is considered.

Figure 10d shows the EBP. The error ‘d’ is used to adapt RPs value to rectify the errors of transmission. It is observed that the errors are rectified by ADSICM during 25×10^3 packets only after the transmission starts. The DICM and ADICM needs more number of packets to send to rectify the errors.

Figure 10e shows the size of HPs required by all the three ICMs. Large header size of 50×10^3 is required by the ADSICM but during the starting phase only i.e. upto 25×10^3 sent packets, which is used to find the error and rectify it. Introduction of integrator in ADSICM reduces EBP further to negligible value, because ADSICM reacts to the increase in the block loss probability by increasing the number of HPs in the sent packets as shown in result. Other two ICM needs HP to be added for more number of sent packets.

The RPs header size has to be adaptive to the CSI received from the receiver. The CSI indicates packet loss probability. If packet loss probability is more, the header size has to be adaptive to this and intern it should increase. When the loss probability is zero, the header size also has to be reduced.

It adds additional parity information among the packets which will reduce the packet loss. The packet loss probability in a critical link is more for DICM, ADICM and reduces drastically for ADSICM. The channel loss probability is kept constant to 0.03 and uniformly distributed. At the receiver side the LPs in a sent packets can be recovered by decoding if no more than the amount of introduced RPs has been lost in that block. As the constant loss probability is introduced in the channel, the sent packets are lost. The number of LPs for DICM is stopped at 48×10^3 sent packets, for ADICM it is stopped at 46×10^3 sent packets, but stopped at significantly lower value of sent packets for ADSICM, typically 28×10^3 as shown in Fig. 10f. For video transmission in wireless end to end delay plays important role in real time transmission. All the three proposed ICMs are evaluated to understand their end to end delay performance and it is observed from Fig. 10g that among all the three suggested schemes, ADSICM has lower end to end delay for the same loss rate. When examining PDR of all the three proposed ICMs, the maximum number of packets are delivered by the ADSICM as shown in Fig. 10h.

As seen in for a loss rate of 0.3 %, DICM gives PDR of 93.0012 %, that for ADICM is 97.9735 % and for ADSICM it is coming to be 99.7974 %. Thus, the algorithm of proposed ADSICM gives maximum PDR when transmitting a video over noisy or unreliable wireless channel.

6 Conclusion

The suggested ADSICM for real-time video transmissions over lossy wireless network can provide a much better end-user experience than existing approaches. The end to end delay is found to be 0.125169 s for ADSICM for the given loss rate of 0.3 % which is much lower. Also, ADSICM automatically modifies the quantity of RPs in FEC encoded packets through CSI by which reduces the number of packets lost significantly by increasing PDR to 99.7974 % for the loss rate of 0.3 %. Hence using ADSICM, transmission errors can be controlled to gain good QoS, so that the stability and good qualities of video transmission can be ensured. Using the suggested ADSICM, it is possible to improve network throughput for reliable transmission of video over noisy and lossy channels. Thus, the suggested ADSICM for real-time video transmissions can provide a much better end-user experience than existing approaches.

References

1. Koli SM, Purandare R, Kshirsagar SP, Gohokar VV (2011) A survey on real time multimedia signal using wireless technology. Adv Netw Commun 132(Part 1):137–147. doi:10.1007/978-3-642-17878-8_15 (Proceedings of first international conference on computer science and information technology, CCSIT 2011, Bangalore, India, 2–4 Jan 2011)
2. Chaari SH, Mnif K, Kamoun L (2012) An overview of quality assessment methods of video transmission over wireless networks. In: Electrotechnical conference (MELECON), 2012 16th IEEE Mediterranean, pp 741–744, 25–28 Mar 2012. doi:10.1109/MELCON.2012.6196537
3. Prasad R, Ruggieri M (2003) Technology trends in wireless communications. Artech House universal personal communications series, London
4. Zhang Q et al (2005) End-to-end QoS for video delivery over wireless Internet. Proc IEEE 93(1):124–234
5. Kuo T-H, Chen P-H, Hung W-C, Huang C-Y, Lee C-H, Yeh P-C (2012) Dynamic source-channel rate-distortion control under time-varying complexity constraint for wireless video transmission. In: Wireless communications and networking conference (WCNC), 2012 IEEE, pp 2566–2570, 1–4 Apr 2012, doi:10.1109/WCNC.2012.6214231
6. Hall JI (2003) Notes on coding theory. Springer, Berlin
7. Yang S, Song W, Zhong Z (2012) Packet-level performance analysis for video traffic over two-hop mobile hotspots. Wireless Commun Lett IEEE 1(2):137–140. doi:10.1109/WCL.2012.022412.110252
8. Issariyakul T, Hossain E (2009) Introduction to network simulator NS 2. Springer science + business media, New York
9. Forouzan BA (2007) Data communications and networking, 3rd edn. McGraw Hill Publication, India
10. Tan Y et al (2008) A real time multimedia signal algorithm over the internet based on FEC and Kalman. In: Proceedings of IEEE international symposium on IT in medicine and education, pp 263–267

11. Aghdasi HS et al (2008) An energy efficient and high quality digital real time multimedia signal architecture in wireless video based sensor networks. Sensors 8:4529–4559. doi:10.3390/s8074529
12. Mansour H et al (2006) Dynamic resource allocation for MGS H.264/AVC real time multimedia signal over link-adaptive networks. IEEE Trans Multimedia 11(8):1478–1491
13. Sohraby K, Minoli D, Znati T (2007) Wireless sensor networks: technology, protocols, and applications. Wiley, New York
14. Schulzrinne H, Casner S (2003) RTP: a transport protocol for real-time applications. RFC3550, The Internet Society
15. Zhang Y et al (2003) Error control and recovery technology in real-time media stream transmission. J Natl Univ Defence Technol 3(6):75–76
16. Angus Lee T-W, Chan G, Zhang Q, Zhu W, Zhang Y-Q (2001) Optimal allocation of packet-level and byte-level FEC in video multicasting over wired and wireless networks. IEEE Globecom'01, Nov 2001
17. Katsaggelos AK et al (2005) Advances in efficient resource allocation for packet-based real-time multimedia signal. Proc IEEE 93(1):135–147
18. Dianat R et al (2006) Reliable real time multimedia signal using codes close to the channel capacity. IEEE Trans Circuits Syst Video Technol 16(12):1550–1556
19. http://www.isi.edu/nsnam/ns/
20. Zeng H, Boyce J (2000) Packet coding schemes for MPEG video over internet and wireless networks. IEEE WCNC, vol 1.3, pp 191–195
21. Ding G, Ghafoor H, Bhargava B (2003) Error resilient real time multimedia signal over wireless networks. IEEE workshop on software technologies for future embedded systems, pp 31–34
22. Harun NZ, Ghazali O (2011) Enhancement on adaptive FEC mechanism for video transmission over burst error wireless network. Information technology in Asia (CITA 11), 2011. 7th international conference on, pp 1–6, 12–13 July 2011. doi:10.1109/CITA.2011.5999532
23. van der Schaar M et al (2003) Adaptive cross-layer protection strategies for robust scalable real time multimedia signal over 802.11 WLANs. IEEE J Sel Areas Commun 21(10):1752–1763

Approximate Minimum Spanning Tree for Points Moving in a Euclidean Two-Dimensions Plane

Anil Kumar Sahu, Chintan Mandal and Suneeta Agarwal

1 Introduction

The Minimum Spanning Tree (MST) is an extensively studied problem and has various graph and geometric based applications. For a connected and undirected graph $G(V, E)$ with positive edge weights, a minimum spanning tree (MST) is an acyclic subgraph of G which connects all the vertices in G such that the total edge weight is minimum. The Euclidean Minimum Spanning Tree (EMST) is a special case of MST where the Euclidean distance between the vertices are taken as edge weight. An EMST is used to setup a minimum cost communication link among networking objects. The objects can be represented as points in the 2D-Euclidean plane. In this paper, we propose an algorithm to maintain an EMST for a set of objects in motion in the 2D-Euclidean plane. This problem was first proposed by Basch [5]. The proposed approach is based on the Kinetic data structure framework [3, 4, 8] which was proposed by Guibas and Basch. Kinetic data structures (KDS) is an event based framework which uses temporal coherence among the involved objects.

Fu and Lee [7] proposed an algorithm for n points in motion which requires an $O(kn^4 \log n)$ preprocessing time, where k is the maximum degree of the motion function of the points. The algorithm requires an $O(m)$ space, where, m is the maximum number of the combinatorial changes of the EMST from time $t = 0\ to\ t = \infty$. After preprocessing, the EMST can be calculated at any given time t in $O(n)$ time.

A. K. Sahu · C. Mandal · S. Agarwal
Department of Computer Science and Engineering,
Motilal Nehru National Institute of Technology, Allahabad, India
e-mail: anilrevolution@gmail.com

C. Mandal
e-mail: chintanmandal@gmail.com

S. Agarwal
e-mail: suneeta@mnnit.ac.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_67,

Agarwal et al. [1] have proposed an algorithm for calculating Kinetic EMST in which the edge weights changes with a linear function of time. Their algorithm runs in $O(n^{2/3}log^{4/3}n)$ per combinatorial change and also supports insertion and deletion of edges. Basch et al. [5] have proposed a data structure to build kinetic EMST which is in $(1+\varepsilon > 1)$ factor of exact EMST, the length of the edge being a function of its end points and ε. Their data structure is local, compact, responsive (which are KDS metrics) and takes $O(\log n)$ time per combinatorial change.
In the KDS framework, a set of certificates are used to track the changes in any high-level attribute of a set of moving objects, e.g. the convex hull of moving points. The certificates are certain geometric relation among the objects which until valid certify the correctness of combinatorial structure of related attribute. Whenever a certificate fails the attribute needs to be updated and the set of certificates must be rebuilt. Failure of each certificate is called an event. Events are classified as external or internal events. External events are those that requires an update in the structure being maintained and also its certificate, e.g. the EMST is a structure that is to be maintained here. Internal events are those that requires updates only in its certificates. The KDS is analysed on the basis of four metrics (responsiveness, efficiency, locality and compactness). A detailed survey about KDS and its analysis can be found in [8].

2 The Proposed Approach

The edges of an EMST for a set of static points in the 2D Euclidean plane is a subset of the edges of the Delaunay triangulation (DT) [6] of same set of points. The EMST can be built in $O(n \log n)$ time by applying Kruskals [9] or Prims [11] over the DT of the points. For points in motion, we can also maintain the EMST while maintaining the DT for the same points. To maintain the DT, we use the KDS to track the changes in the Kinetic Delaunay Triangulation (KDT) proposed by Guibas et al. [2].
Let $S = \{P_1, P_2, \ldots, P_n\}$ is the set of n points moving in the 2D-Euclidean plane. Each point P_i is represented by its motion equation,

$$P_i = (x_0 + v_x * t, y_0 + v_y * t), P_i \in S \tag{1}$$

x_0, y_0 is the coordinates of the point at time $t_0 = 0$ and v_x, v_y gives the velocity of the point along X and Y axis. We want to maintain the EMST for the points in S. We assume that the points are moving without collision. This assumption is required by the KDT on which our solution is based. Initially, at t_0, the EMST is built over the DT of S. All edges of DT which are in EMST are referred to as *branch* and non-EMST edges as *chord*. We constrain the points to move in a rectangular boundary, bounded by four points at infinity. Thus the convex hull of the DT will always be this rectangular boundary. However, if we do not make this assumption, then the convex hull also has to maintained, which is the boundary of the DT. The convex hull can be maintained with the help of two hull certificates. One hull certificate is used to track the changes in convexity of the boundary and the other to check that whether the triangles at boundary are properly oriented or not [12].

The change in EMST depends on the edge lengths and as the points are moving, edge lengths will change continuously. The EMST may also change whenever structure of the DT changes. This happens when an incircle of a triangle in the triangulation contains another point of S. This failure effects a flip of the common edge of the quadrilateral formed by the triangle and the point lying on the incircle. If the flipped edge in question is a chord, it is referred to as *flip event* [2]. If the flipped edge is a branch, we will refer to it as an *Effective flip event*.

As the points are in motion, it is possible that no flip event occurs for a period of time and the DT does not change for a long time even though the length of the edges of the DT changes. This creates a difference in the total cost of the EMST and the exact EMST of the points at that time. Therefore, to keep the EMSTs cost as near as possible to that of the exact EMST we consider a parameter *Stretch factor* $\lambda (0 < \lambda < 1)$. Using λ, we create a certificate which we refer to as *Stretch Certificate*.

$$length(E_i)_t \leq (1 + \lambda) * length(E_i) \; at \; t_{cur} \tag{2}$$

where, $length(E_i)_t$ is the timed function of the squared Euclidean length of edge E_i and $length(E_i)$ is squared length of E_i at current time. It checks whether the length of any branch is within $(1 + \lambda)$ factor of its length at current time. The EMST may change on the failure of this certificate which we refer to as *Stretch Event*.

2.1 Handling Events

2.1.1 Flip Event

Whenever an incircle certificate fails, it is referred as a *flip event*. This event is handled by flipping the edges and calculating five new incircle certificates and their failure time for all the edges in adjacent triangles of the flipped edge.

2.1.2 Effective Flip Event

An *Effective flip event* occurs when a *branch* is flipped. The *Effective flip event* is handled by searching locally for a potential edge that can reconnect the EMST. By local searching, we limit the search amongst the chords forming the triangle adjacent to the flipped edge. Out of all the candidate edges we choose one that is shortest and does not create a cycle. A check for a cycle in the graph is done in $O(1)$ time by checking the adjacent two edges to be branches of the triangle incident on the chord in question. The check can also be done by a breadth first search along the branches

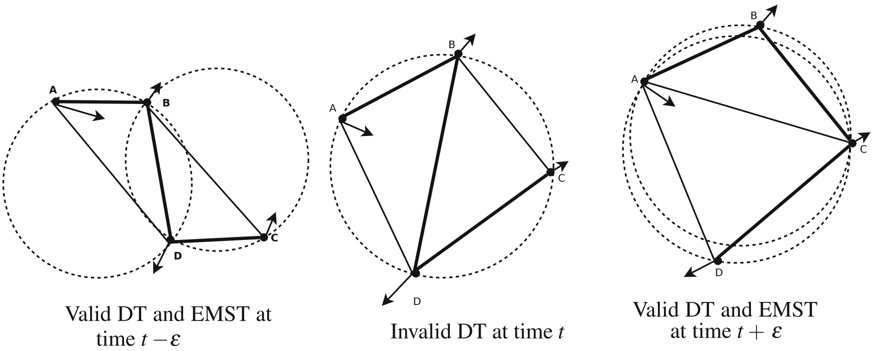

A
B
C
D
Valid DT and EMST at time $t-\varepsilon$
Invalid DT at time t
Valid DT and EMST at time $t+\varepsilon$

Fig. 1 Effective flip event

of the DT, which requires $O(n)$ time. The time required to update the event queue is $O(n \log n)$ time if the check for a cycle is done by breadth first search else it requires $O(\log n)$.

Figure 1a shows a valid DT and EMST of four points A, B, $C\&D$ at some time $t - \varepsilon$. At time t Fig. 1b, the DT becomes invalid which is updated by flipping the branch **BD**. Before flipping the branch **BD**, a new branch is searched to keep the EMST connected. As $BC < AD$ is considered as the new branch, **BD** is flipped with **AC** Fig. 1c.

2.1.3 Stretch Event

When a stretch certificate fails, we search locally for the presence of a chord which is shorter than the edges and that does not create cycle. If a potential chord is found, we update the EMST by interchanging chord and branch. For a *Stretch event*, we check the presence of a cycle by applying breadth first search along the EMST. This event can be processed in $O(n)$ time. In Fig. 2, at t, the stretch certificate for edge **CD** fails. A potential chord is searched in the $\triangle ACD$. Since $AD < \{AB, BC, CD\}$ is the shortest amongst other chords locally and also the branch under consideration, the stretch certificate for **CD** fails and there is an updation of the EMST.

2.2 *Our Algorithm*

The algorithm shows the way to construct initial EMST from DT and the steps required to maintain the EMST.

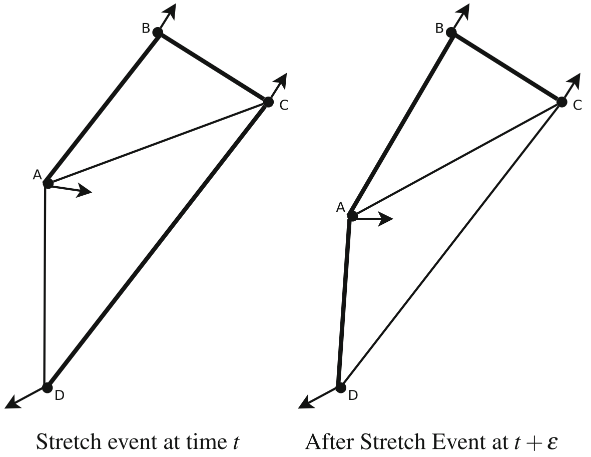

B
C
A
D
B
C
A
D
Stretch event at time t
After Stretch Event at $t+\varepsilon$

Fig. 2 Stretch event

Algorithm Construction and Maintenance of EMST

```
Input: $S = \{P_0, P_1, \ldots, P_n\}$ : Set of $n$ points; $Q$: Priority Queue
Output: Continually updated Approximate EMST of points in $S$.
$Ev$: Event variable; $t_{cur}$: current time
/**********Initialization Step($t_0 = 0$)**********/
Create EMST from Delaunay triangulation, $DT(S)$ of points at time $t_{cur}$.
for $\forall$ Edges e $\in DT(S)$ do
  if e is a common edge then
    Calculate incircle certificate
  else if e is an MST edge then
    Calculate stretch certificate
  end if
end for
for Each Certificate $C$ do
  Calculate failure time and store it in $Q$.
end for
/**********Iteration Step**********/
while points are in motion do
  Advance Time t
  if $t = TimeinQ.head()$ then
    $Ev \leftarrow Q.pop()$
    Process $Ev$
    Push new events into $Q$ as required
  end if
end while
```

The algorithm starts by constructing the Delaunay triangulation for the points in S at initial time $t_0 = 0$ and find the EMST on $DT(S)$. Create all the certificates and calculate the failure time of all the certificates and store them as events in the priority queue, Q. Head of Q will always point to the event with smallest failure time. In step-2, we advance time t, and keep on checking whether the current time is equal to the time of the event of head of Q. When current time reaches the time of the first event, the event is popped from Q. The event is processed and new events are inserted as required into the Q.

2.3 Performance Analysis

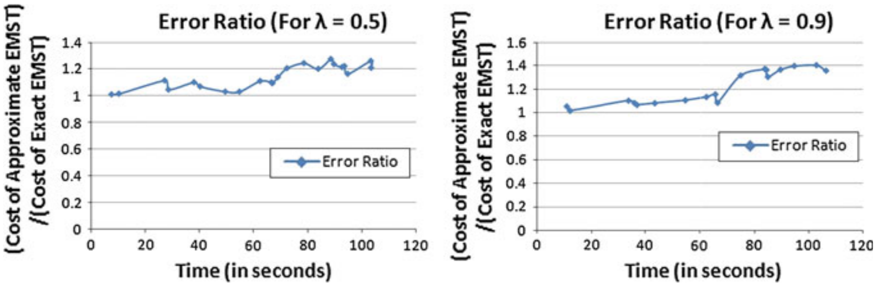
The EMST from DT can be constructed in $O(n \log n)$ for static points as preprocessing time. Since, total number of edges in DT is $O(n)$ and total number branches is $(n - 1)$, all the certificates (incircle and stretch) and their failure time can be calculated in $O(n)$ time. In accordance to KDS metrics,

1. The total number of DT edges is in $O(n)$. The total number of certificates is also in $O(n)$ and each point participates in average of six (average degree of a vertex in DT), i.e. $O(1)$ number of certificates. Thus, it is local and compact.
2. To process flip event and effective flip event it takes $O(\log n)$ which is the time required to update the event queue. But, it requires $O(n)$ time to process *Stretch event* because it uses breadth first search to detect a cycle. So, in overall our method is not responsive.
3. It is not possible to give upper bound that how many stretch event turns out to be external i.e. requires update in EMST. So, we do not analyze whether our approach is efficient or not.

The EMST obtained is approximate, as the branches chosen are locally chosen in case of the external events. A spanning tree can be transformed from one tree to another by exchanges of chords in the graph [10]. The approximate EMST obtained can be exchanged with $(n - 1)$ chords to obtain the exact EMST. Thus, the approximate EMST is $O(n)$ approximate of the exact EMST.

3 Experimental Results

We have implemented our algorithm in Java and JOGL (Java Bindings for Open GL). We have done experiment for points moving with linear velocity for various stretch factors. We give here results for $\lambda = 0.5$ and $\lambda = 0.9$. The error ratio, which is the ratio of the cost of length of the approximate EMST and the length of the exact EMST (computed using Kruskals algorithm on the points considered as static) is computed. The comparison is done whenever there is any change in the approximate EMST. Figure 3a, b shows the variation in error ratio from time t $=$ 0 to t $=$ 115 seconds for 10 points.The error ratio obtained is less than $1 + \lambda$.



Variation in error ratio for λ = 0.5

Variation in error ratio for λ = 0.9

Fig. 3 Variation in error ratio for 10 points

4 Conclusion

We have proposed an algorithm for maintaining an approximate EMST for points moving in 2D-Euclidean plane. Although, efficiency and responsiveness are more important metrics of KDS, but our approach is only local and compact. We can make it responsive by using some extra space for storing the approximate EMST in such a way, such that the cycle can be detected in less than $O(n)$ time. The datastructure used in our algorithm do not require any central storage. It performs in the same way if it is distributed over the points. Thus, this algorithm can also be used to find an EMST for independent objects in motion having autonomous processing capabilities.

References

1. Agarwal PK, Eppstein D, Guibas LJ, Henzinger MR et al (1998) Parametric and kinetic minimum spanning trees. In: 39th annual symposium on foundations of computer science, pp 596–605
2. Albers G, Mitchell JSB, Guibas LJ, Roos T et al (1998) Voronoi diagrams of moving points. Int J Comput Geom Appl 8:365–380
3. Basch J (1999) Kinetic data structures. Ph.D. thesis, Stanford University
4. Basch J, Guibas LJ, Hershberger J et al (1997) Data structures for mobile data. In: Proceedings of the eighth annual ACM-SIAM symposium on discrete algorithms, SODA '97. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, pp 747–756
5. Basch J, Guibas LJ, Zhang L et al (1997) Proximity problems on moving points. In: Proceedings of the thirteenth annual symposium on Computational geometry, SCG '97. ACM, New York, NY, USA, pp 344–351
6. Devadoss SL, O'Rourke J et al (2011) Discrete and computational geometry. Princeton University Press, Princeton
7. Fu J-J, Lee RCT et al (1991) Minimum spanning trees of moving points in the plane. IEEE Trans Comput 40(1):113–118
8. Guibas LJ (1998) Kinetic data structures—a state of the art report
9. Kruskal JB (1956) On the shortest spanning subtree of a graph and the traveling salesman problem. Proc Am Math Soc 7:48–50

10. Mare M (2008) The saga of minimum spanning trees. Comput Sci Rev 2(3):165–221
11. Prim RC (1957) Shortest connection networks and some generalizations. Bell Syst Tech J 36:1389–1401
12. Russel D (2007) Kinetic data structures in practice. Ph.D. thesis, Stanford University

Low Complexity Speech Enhancement Algorithm for Improved Perception in Mobile Devices

B. S. Premananda and B. V. Uma

Abstract In mobile phones, perceived quality of speech signal deteriorates significantly in the presence of background noise since near-end/surrounding noise also arrives at the near-end listener's ears. The quality of the received signal varies widely depending upon signal strength and unavoidable background noise in the user environment. There is a need to improve the quality of received speech signal in noisy conditions by developing the speech enhancement algorithms. This paper focuses on the impact of the various background noises on signal perception and mechanisms to mitigate the noise impact for improved signal perception. Gain adjustment process with simple time domain approach has been adapted to improve the quality and intelligibility of the speech signal in the noisy environments by automatically increasing output levels when the noise dominates. Since time domain approach has been used, it is less complex and consumes very less battery power of the mobile and also efficiently masks the background noise. This paper studies the effect of various signal parameters on automatic gain adjustment of the received signal to enhance its perception.

Keywords Adaptive speech enhancement · Background noise · Degradation · Gain · Perception.

B. S. Premananda (✉)
Department of Telecommunication Engineering, R.V. College of Engineering, Bangalore, India
e-mail: premanandabs@rvce.edu.in

B. V. Uma
Department of Electronics and Communication Engineering, R.V. College of Engineering, Bangalore, India
e-mail: umabv@rvce.edu.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_68,

1 Introduction

Mobile phones are essential component of day to day life. Day by day they are becoming very affordable making the communication infrastructure as the backbone of daily life. Background noise is a natural part of a conversation. In most cases, the background noise of the environment in which the source of speech lies, is the main component of noise which adds to the speech signal. A high background acoustic noise level is annoying to the listener side. Listener fatigue (the ears get tired) and difficult to understand each other. Considerable attention has been paid over the past decade for the enhancement of speech degraded by additive background noise. Listening to speech or audio signals becomes more difficult as the background noise level dominates. Hence, there is a strong need to improve the quality of the speech signal in noisy conditions by developing speech enhancement algorithms to minimize the effect of background noise. In the gain adjustment process the quality of the speech signal in noisy environment is improved by automatically adjusting the output level when the background noise exceeds the noise masking threshold. The traditional noise cancellation algorithms cannot be used as the near-end noise directly arrives at the near-end listener's ears. The approaches proposed for far-end noise cancellation are discussed in literature [1–3]. Several approaches to mitigate the background (near-end) noise using speech enhancement are discussed by Bastian et al. [4–6] and Jong Won Shin et al. [7, 8]. Speech intelligibility improvement in presence of near-end noise and loud-speaker output power constraint are discussed in [4, 5] and [6]. Perceptual speech reinforcement based on partial specific loudness is discussed in [7, 8].

2 Proposed Method

This paper proposes a method to overcome the relative degradation of the speech signal in the noisy environment. In order to retain the signal quality in the noisy environment different algorithms have been proposed and implemented in [4, 7, 8], but the complexity involved in determining the gain seems to be unresolved. In this paper a concept which is simple to analyze and implement in real time is presented. A multiplier is required to enhance the speech signal degraded in the presence of background noise. As the speech signal is dynamic it is difficult to find a constant multiplication factor (MF) for the incoming signal. Thus the signal strength and noise present in the environment is to be analyzed. The speech signal is to be amplified by applying a dynamically varying gain depending upon the different parameters of speech and noise signal. In order to avoid the perceptual artifacts like clicks & pops due to signal saturation/overflows and signal bursts due to sudden gain changes, optimal gain is computed which is characterized by slow and configurable response time for the gain variations. Block diagram of the proposed approach is shown in the Fig. 1. Energy of the downlink speech signal and near-end noise signal is computed.

By comparing energy of both the speech and noise signals, gain/MF is computed for enhancing speech signal. Gain obtained is multiplied with speech (element by element) signal to get enhanced speech signal. The external volume control of the devices can't be used for this purpose as it is painful.
The effective gain/multiplication factor depends on the following parameters:
i. **Energy**: Sound energy, energy produced by the sound vibrations as they travel through a specific medium. The sound energy is proportional to the square of the amplitude. Energy, E_{dB} in decibels is calculated using,

$$E_{dB} = 10 * \log_{10} \sum (X^2) / N \tag{1}$$

where X is the amplitude of the speech signal and N is the number of samples.

ii. **Intensity**: Intensity of sound waves is defined as the average energy transported per second per unit area perpendicular to the direction of propagation. It is measured in $Js^{-1}m^{-2}$ or Wm^2.

iii. **Loudness**: Loudness of sound is the degree of sensation of sound produced in the ear. Loudness of sound depends on its intensity but the relationship is not linear. The magnitude of sensation is proportional to the logarithm of the physical stimulus which produces it.

iv. **Type of Noise**: In hammering type noise, attack and release concept is used. This indicates that when the hammer like noise is active, the actual signal is to be amplified immediately and stay at the amplified level till the noise diminishes. After the noise diminishes the amplified signal is to fall down slowly not suddenly, which in effect is similar to a low pass filtering is reflected, the amplification shoots up then stay at the appropriate point and is released slowly.

v. **User Perception Level**: No standalone technique can be efficient for all kinds of users. Different users have different perception levels and abilities to distinguish signal and noise signal. Hence, a user factor also needs to be incorporated while calculating the multiplication factor.
The Gain/MF is to be adaptive for being effective in most of the scenarios. It can be modeled as,

$$\text{Gain} = \sum_{i=1}^{p} c_i x_i \tag{2}$$

where $x_1 \ldots x_i \ldots x_p$ represent different parameters and $c_1 \ldots c_i \ldots c_p$ represent different coefficients to be derived corresponding to one or more of the above mentioned parameters.

The challenge lies in finding out the correct coefficients for adapting Eq. (2). Hit and trial methods or human expertise can be utilized for setting these values, works for limited categories of speech signals and noises.

3 Deriving the Gain

The signal and noise parameters discussed above can be considered as distinct features for deriving the gain. The correct gain/MF for a pair of signal and noise sample (selected parameters) is user specific. It is assumed that the target system provides a mechanism for user to declare whether system generated MF (auto gain) is acceptable or not. Such an assumption is easy to realize for most of the devices as volume control can be used for the purpose. In Fig. 1, Gain block calculates the amplification factor/MF of speech signal in the presence of varying background noise which can dominate the received speech signal as well. Different parameters of signal and noise (energy in the present case) can be considered.

The Gain is derived using the equation,

$$\mathrm{Gain} = \mathrm{A} + \mathrm{maximum}\ (\mathrm{B}, (\mathrm{C} - \mathrm{D})) \cdot \mathrm{E} \tag{3}$$

where A, B, C, D, and E are experimental constants. Value of A (default gain) is set to 1 so that Gain = 1 when no enhancement for the speech signal is required, B (default enhancement) is set to 0. (C – D) is extent of noise over speech signal, (if negative no amplification is required), E (compensation factor) is used to control the gain (<1).

If the hearing and perception choice of user varies with time, the proposed mechanism is capable of capturing such alterations. If the amplification value exceeds the maximum loudness [5] of the loud speaker [10] then end capping can be performed depending on minimum and maximum values computed using,

Case-1: for positive amplified value (AV)

$$\mathrm{AV} = \mathrm{minimum}(\mathrm{AV}, 2^{15} - 1) \tag{4}$$

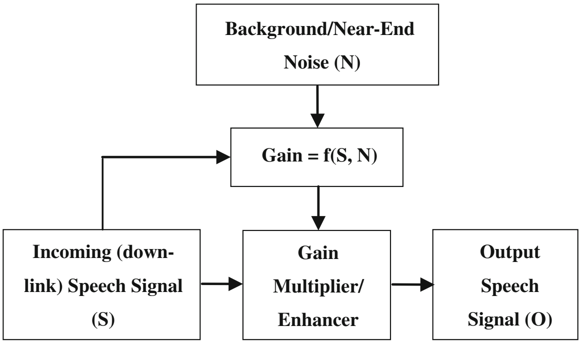

Background/Near-End Noise (N)
Gain = f(S, N)
Incoming (down-link) Speech Signal (S)
Gain Multiplier/ Enhancer
Output Speech Signal (O)

Fig. 1 Proposed approach

Case-2: for negative amplified value (AV)

$$\mathrm{AV} = \mathrm{maximum}(\mathrm{AV}, -2^{15}) \tag{5}$$

Signal buffer value can be overridden with new amplified value and output to output signal buffer.

4 Experimental results

Algorithm has been verified for different speech signals with varying background (tapering, hammer type ...) noise signals. To **illustrate,** a speech signal of 16 bit PCM with sampling rate of 8 KHz with duration of 5 s is considered. The speech signal is captured using an audio editor tool Gold Wave and saved in .wav format. The captured speech signal has 40000 (8000*5) samples, total samples are divided into frame size of 1024 each, resulting in 39 frames. Varying noise sampled at 8 KHz is considered for experimenting and gain/MF are calculated.

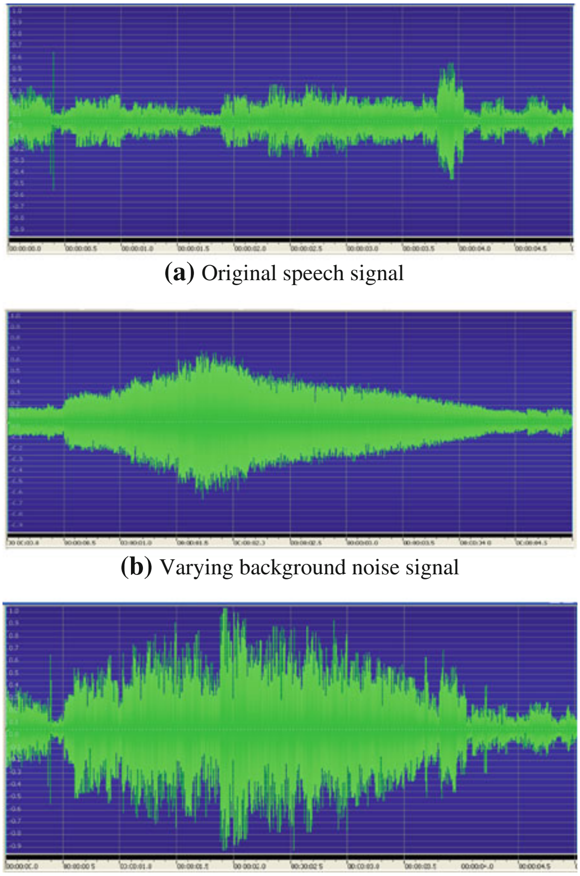
Figure 2 show the results obtained for varying background noise. Test is performed by considering the difference in energy of background noise (dB) and speech signal (dB).

The speech signal in Fig. 2a corresponds to input (downlink) speech signal, Fig. 2b represents noise signal whose amplitude is continuously decreasing, Fig. 2c shows the amplified speech signal in presence of noise where the gain is varied from 1 to 6.8156, values are depicted in Table 1. Figure 2c reflects the variation in the amplification of speech signal based on changing noise signals. Table 1, reflect the values of gain/MF for the varying speech and noise signals.

Enhanced speech (energy) signal is given in last column (new SE dB). Speech signal can be further enhanced by increasing the value of gain control, E in Eq. 3 as desired. When signal energy (SE) is sufficiently greater than noise energy (NE) then gain of 1 is used so that no amplification is required, input signal buffer value will be copied to output signal buffer. When SE is approximately equal to NE then gain is selected such that enhanced speech signal is increased nearly by 1 dB. When SE is less than NE then gain is calculated using Eq. 3. Present frame gain is compared with previous frame gain if the difference is more, then current frame gain is adjusted to avoid sudden change in large gain to avoid click and pop noise.

The obtained results are plotted in Fig. 3. Number of frames is indicated in x-axis and y-axis indicates energy of the speech, noise & enhanced signal and variation of gain w.r.t. speech and noise signal. For a gain of 6.8156 the speech signal energy is enhanced by 16.67 dB. Speech signal has enhanced to a maximum energy of 81.6397 dB when the noise signal is at maximum.

The time domain approach has been adapted to obtain the gain, no frequency transformations were used hence less number of multipliers, adders and load & store instructions were required for processing the algorithm. The proposed algorithm



(a) Original speech signal

(b) Varying background noise signal

(c) Enhanced speech signal

Fig. 2 **a** Original speech signal. **b** Varying background noise signal. **c** Enhanced speech signal

requires 1000+ multipliers, adders, and load & store operation for processing one frame (1024 samples) of the signal. Although the MF obtained are close to transformed method using absolute threshold of hearing, in terms of complexity and power consumption the proposed approach is suitable for enhancing the speech signal in real time. Since Battery is critical for mobile devices the proposed algorithm consumes very less battery power for processing. Limitation of the algorithm is that amplification of the overall power of the speech signal according to the noise level. Simulation results were verified using an audio editor tool Gold Wave v5.58, compiled with Eclipse C/C++ IDE and implemented on Beagle-Board-xM (OMAP-3530) platform.

Table 1 Derived gain for speech and noise signal energy

Frame no.	SE dB	NE dB	Gain/MF	New SE dB
01	72.5013	67.3345	1	72.5013
02	71.0811	67.1478	1	71.0811
03	69.7505	67.5765	1	69.7505
04	67.9069	67.0283	1	67.9069
05	67.8812	70.6206	1.82182	73.0906
06	69.0009	73.295	2.28822	76.1903
0 7	68.5412	73.128	2.37603	76.0577
08	68.201	73.7001	2.64972	76.6645
09	65.1991	75.5899	4.11722	77.4908
10	65.641	76.5888	4.28434	78.2782
11	65.7944	77.7956	4.60035	79.0499
12	64.2722	78.1714	5.16976	78.5412
13	63.9847	79.2568	5.58163	78.9195
14	61.1355	80.5208	6.81561	77.8051
15	64.3364	80.2096	5.76196	79.501
16	69.3247	79.7831	4.13751	81.6397
17	69.1349	78.8431	3.91247	80.9836
18	66.9738	77.3582	4.11532	79.2615
19	71.6405	77.1728	2.65969	80.1367
20	69.3675	76.5438	3.15289	79.3407
21	70.6711	76.2115	2.66213	79.1753
22	70.4634	75.5076	2.51328	78.4678
23	69.0819	75.1582	2.8229	78.0953
24	69.7312	75.2283	2.64914	78.1928
25	68.2743	74.8407	2.96991	77.7287
26	68.4736	74.4761	2.80076	77.4186
27	68.9679	73.7095	2.42248	76.6526
28	67.9661	72.9919	2.50775	75.9512
29	67.319	72.1953	2.46287	75.1473
30	70.1954	71.4764	1.38431	73.0193
31	76.0615	70.347	1	76.0615
32	70.7926	69.2804	1	70.7926
33	64.0397	67.9061	2.15993	70.7275
34	67.862	65.841	1	67.862
35	66.3948	65.4736	1	66.3948
36	63.0688	64.1974	1.33858	65.5998
37	68.3801	65.4015	1	68.3801
38	65.7227	64.9955	1	65.7227
39	63.8682	64.6441	1.23278	65.6841

SE Signal Energy, *NE* Noise Energy

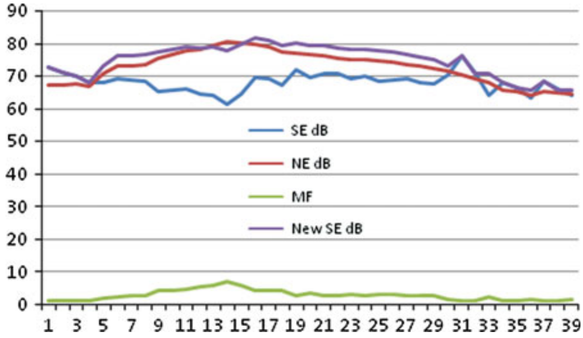


Fig. 3 Speech, noise and enhanced signal with MF noise

5 Conclusions

The paper discusses the implication of contextual background noise on signal perception by the user and necessity of this problem needs to be solved for extended user experience and growth of mobile technology. The paper also discusses avenues to mitigate the effect of such noise by gain adjustment of speech signal with varying noise signal. Experimental results are presented for a varying background noise, highlighting acceptable gain of the signal. Results indicate that gain obtained is adaptive and varies w.r.t. change in speech and noise signal. Proposed system is capable of tracking changes in user perception levels with time.

References

1. Boll SF (1979) Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans Acoust Speech Signal Process ASSP-27:113–120
2. Ephraim Y, Malah D (1984) Speech enhancement using a minimum mean square error short-time spectral amplitude estimator. IEEE Trans Acoust Speech Signal Process ASSP-32:1109–1121
3. Virag N (1999) Single channel speech enhancement based on masking properties of human auditory system. IEEE Trans Speech Audio Process 7:126–137
4. Sauert B, Vary P (2006) Near end listening enhancement: speech intelligibility improvement in noisy environments, In: Proceedings of international conference on acoustics, speech, and, signal processing (ICASSP), pp 493–496
5. Sauert B, Enzner G, Vary P (2006) Near end listening enhancement with strict loudspeaker output power constraining. In: Proceedings of international workshop on acoustic echo and noise, control (IWAENC)
6. Sauert B, Vary P (2009) Near end listening enhancement optimized with respect to speech intelligibility index. In: Proceedings of european signal processing Conference (EUSIPCO) 17. EURASIP. Hindawi, New York, pp 1844–1848

7. Shin JW, Kim NS (2007) Perceptual reinforcement of the speech signal based on partial specific loudness. IEEE signal process lett 14(11):887–890
8. Shin JW et al (2007) Speech reinforcement based on partial specific loudness. In: Proceedings of European conference on speech communication an technology (EUROSPEECH), pp 978–981

Prolonging the Lifetime of Wireless Sensor Network by Exponential Node Distribution and Ant-colony Optimization Routing

Zaheeruddin, Aruna Pathak and Manoj Kumar Tiwari

Abstract Wireless Sensor Networks (WSN) consists of small nodes to collect useful information from environment, but nodes have limited power. Nodes in WSN transmit data to the sink which is fully equipped with energy source and from sink data is transmitted to base station which is capable of connecting the WSN to an existing communications infrastructure or to Internet where user can access to reported information. Sensor nodes near to sink is transmitted their own information as well as forward information from the outer nodes. After some time, sensor nodes near to the sink will lose their energy and then die although outer region nodes are still alive, but information cannot be transmitted to sink because inner region sensor nodes are dead means energy hole comes near to sink, so lifetime of WSN is affected. To minimize energy-hole problem near sink and prolong lifetime of WSN, different approaches are used. In our approach, nodes according to exponential distribution are deployed then ant colony optimization technique is used to route the information.

Keywords Energy hole · Exponential distribution · Ant colony optimization routing · Wireless sensor network

Zaheeruddin (✉) · A. Pathak
Department of Electrical Engineering, Jamia Millia Islamia Central University,
New Delhi, India
e-mail: zaheeruddin@jmi.ac.in

A. Pathak
e-mail: aruna.pathak@gmail.com

M. K. Tiwari
TR&D, STMicrolectronics Pvt. Ltd, Greater Noida, India
e-mail: mkt.unnao@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_69,

1 Introduction

Wireless Sensor Network (WSN) consist of small nodes with sensing, computation, and wireless communications capabilities [1]. Basically WSNs contain hundreds or thousands of sensor nodes, and these sensors have the ability to communicate either among each other or directly to an external base station (BS). A WSNs applications may include military applications, environmental applications, health applications and other commercial applications [2]. Basically nodes are driven by batteries and in many applications it is not easy to replace the batteries or sometimes not even recharge the batteries so each node has a limited energy supply. Lifetime of a sensor network may be defined as the time till the first node dies. Although the time until the first node fails is an important measure from the complete network coverage point of view, this performance metric alone cannot measure the lifetime performance behaviour for all nodes in the network. Information collected at the remaining nodes can still be delivered successfully to the base-station. So lifetime may also be defined as the time till a proportion of nodes die. The location of the failure nodes is also of importance. If the proportions of nodes that have run out of battery are located in some critical part of the network, e.g., connecting the central sink and the rest of the network, it may result in early dysfunction of the entire network. Our discussion will be taken in the spirit of the second definition. Now we come to an important point related to lifetime which is energy hole problem. Sensor nodes sitting around the sink need to relay more traffic compared to other nodes in outer sub-regions, nodes in inner regions suffer much faster energy consumption rates and thus have much shorter expected lifetime. This phenomenon of uneven energy consumption rates is termed as energy hole problem, which may result early dysfunction of the entire network, even if the other parts of the network still have a lot of energy. Energy hole problem proved by Olariu and Stojmenovic [3]. They have done an experiment under the uniformly distributed network. They have divided the experiment area into concentric coronas from the sink. And Simulation results show that, when the first corona consumed all the energy, the energy expended by a sensor in the 10-th corona is only about 4.197 %. Li and Mohapatra investigate the problem of uneven energy consumption in a large class of many-to-one sensor networks [4, 5]. They observed that, hierarchical deployment and traffic compression could have positive effects on the problem, while heavy node density or more data generation have little or negative effects on it. Non-uniform distribution method [6–8] is proposed to solve the energy hole problem. It has been demonstrated that non-uniform distribution method can provide a balanced energy consumption ratio of the nodes. In this paper we first deploy the nodes according to exponential distribution, and then we use optimal transmission range so that the lifetime of whole network will be improved. Simulations results show that we can achieve more energy efficiency than uniform distribution and Geometric distribution. The remainder of the paper is organized as follows: In Sect. 2 we present the survey on the related work. As the basic models and assumptions are described in Sect. 3, 4 proposes the exponential distribution of nodes and ant colony optimization routing. Section 5 describes the simulation results of the proposed schemes. Finally, Sect. 6 concludes the paper and describes the future work.

2 Related Work

Energy hole problem first investigated by Olariu and Stojmenovic. They propose a corona width model under uniform node distribution [3]. The network area is divided into concentric coronas from the sink. They use corona different width to balance the node energy dissipation. As a result, the outer coronas have larger corona width than the inner coronas, and the energy consumption can be expended evenly. However, this model cannot achieve an energy efficient consumption because most nodes in the model cannot get an optimal transmission route to sink. Li and Mohapatra [4] investigate the problem of uneven energy consumption in a large class of many-to-one sensor networks. The authors describe the energy hole in a ring model (like corona model), and present the definitions of the per node traffic load and the per node energy consuming rate. Based on the observation that sensor nodes sitting around the sink need to relay more traffic compared to other nodes in outer sub-regions, their analysis verifies that nodes in inner rings suffer much faster energy consumption rates and thus have much shorter lifetime. The authors term this phenomenon of uneven energy consumption rates as the energy hole problem, which may result in serious consequences, e.g. early dysfunction of the entire network. Shiue [9] propose an algorithm to resolve energy hole problem, which uses mobile sensors to heal energy holes. But, the cost of these assistant approaches is a lot. Wu propose a non-uniform node distribution strategy which can achieve a sub-balanced energy consumption of nodes. Authors state that if number of nodes from outer corona to inner corona increases by an exponent q. The total number of nodes in the network would be $N = NR \cdot qR - 1$ (R represents the number of coronas, NR represents the node number in the out most corona). But it is not practically worked because the model assumes that each sensor node should generate data with same length at a unit time. The consequence is that: for the same size of areas, the area near the sink generates more data than the area away from the sink in the same time. In this paper we jointly work on non-uniform node distribution strategy i.e. positive exponential distribution towards the sink, which can achieve a sub-balanced energy consumption of nodes and optimal transmission range so that the lifetime of wireless sensor network will be improved.

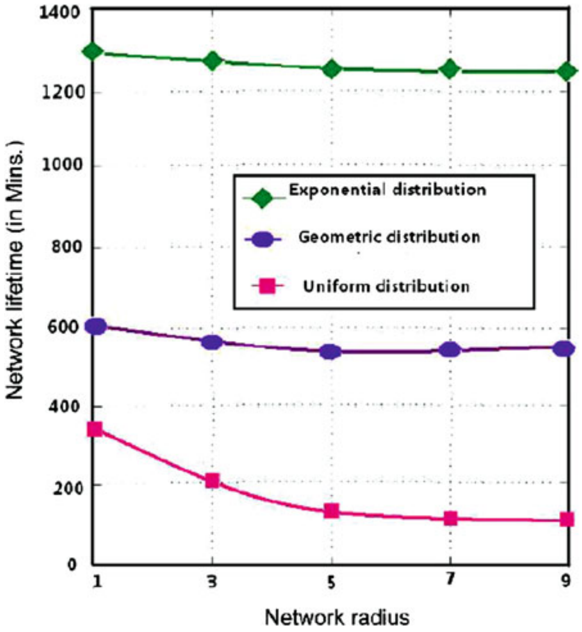


Fig. 1 Network lifetime of different node distributions

3 Preliminaries

3.1 Network Model

We assume that a set of homogeneous nodes are deployed in a circular area of radius R. The sink is located at the center of the circle, dividing its sensing fields into 9 concentric coronas named as C1, C2 ...C9. All sensor nodes are static, each of which has knowledge of its own position and the position of the sink. Nodes generate data at the same constant rate. We assume a MAC layer which guarantees there are no collision and retransmission. From the viewpoint of the whole network, the nodes are distributed exponentially. Therefore the density are related as follows

$$\rho 1 > \rho 2 > \rho 3 > \rho 4 > \ldots\ldots\ldots\ldots > \rho 9 \tag{1}$$

In Fig. 1, a darker corona shows a higher node density.

3.2 Energy Model

In our model, only the energy of data transmission and data reception consume energy. We share the same idea with [8] that sampling and idle listening consume limited energy. Each sensor has an initial energy ε. The energy cost of data transmission Et and the energy cost of data reception Er are defined as the Eq. (1) and the Eq. (2), where dt represents the transmission distance between the source and destination, L represents the data length of the transmission by bits.

$$\mathrm{Et} = \beta\,1 + \beta\,2\ \mathrm{dt}\,\alpha\,\mathrm{L} \tag{2}$$

$$\mathrm{Er} = \beta\,3\mathrm{L} \tag{3}$$

According to [4], the values of parameter β 1, β 2, β 3 are set as follows. $\beta\,1 = 45 \times 10 - 9\mathrm{J/bit}$, $\beta\,2 = 0.001 \times 10 - 12\mathrm{J/bit/m4}$, $\alpha = 4$, $\beta\,3 = 135 \times 10 - 9\mathrm{J/bit}$The dt α accounts for the path loss, and we let $\alpha = 4$ in this paper.

4 Jointly Approach of Exponential Node Distribution and Ant Colony Optimization Routing

In this section we present the jointly effect of exponential node distribution toward sink and ant colony optimization routing on the lifetime of wireless sensor network.

4.1 Exponential Node Distribution

As we know that Sensor nodes near the sink have to transmit more data compared to other nodes in outer sub-regions, so that nodes which are near the sink lose their energy very quickly and thus have much shorter expected lifetime. This phenomenon of uneven energy consumption rates is termed as energy hole problem, which may result early dysfunction of the entire network, even if the other parts of the network still have a lot of energy. Our first approach to reduce this problem with the help of exponential distribution of nodes towards the sink. A random variable x is said to follow the exponential distribution with parameter λ if its distribution function F is given by: $F(x, \lambda) = 1 - \exp(-\lambda x)$ for $x \geq 0$

$$0 \quad \text{for} \quad x < 0 \tag{4}$$

Let $F(x, \lambda) = Y$
Then $\quad Y = 1 - \exp(-\lambda x)$

$$
\begin{aligned}
&1 - \mathrm{Y} = \exp(-\lambda \mathrm{x}) \\
&\ln(1 - \mathrm{Y}) = \ln\{\exp(-\lambda \mathrm{x})\} \\
&\ln(1 - \mathrm{Y}) = -\lambda \mathrm{x} \\
&\mathrm{x} = -1/\lambda\{\ln(1 - \mathrm{Y})\}
\end{aligned}
\tag{5}
$$

4.2 Ant Colony Optimization Routing

Ant colony optimization routing randomly select a search path. It does not need any prior information at first. Ant colony optimization routing includes three main aspects-

(i) Pheromones are the media and ants communicate to each other with the pheromones. When the associate selects the path, ants will choice paths according to the pheromone on the path.
(ii) The paths searched by the ants can be committed to the memory, the paths will not be selected in the next search, so the taboo list is created in algorithm simulations.
(iii) If number of ants through some paths are few, then the pheromone will be evaporation in the paths over time. If the number of ants through some paths is large, the number of the pheromone will be more and more. Such that the information intensity will increase, the probability of choosing this path will increase and the intensity of the pheromone in this path will be further greater. So we can simulate the phenomenon to establish the route choice mechanism and make the search of the ant colony algorithm towards the optimal solution. The search mechanism of the ant colony algorithm displays a positive feedback or autocatalytic characteristics. Each ant tries to find a path in the wireless sensor network with minimum cost. Ants start from the node which have the data for transmission called source node ‘S’ and then move through neighbour nodes N*i*, and reach to finally sink node(destination) D. Choice for the next node N is done according to the probabilistic decision rule (6):

$$
\mathrm{P_k(N, S)} = \begin{cases} \dfrac{[\delta(\mathrm{N, S})^{\mu}][\varphi(\mathrm{N, S})^{\tau}]}{\sum\limits_{\mathrm{NERs}} [\delta(\mathrm{N, S})^{\mu}][\varphi(\mathrm{N, S})^{\tau}]} & \text{if } \mathrm{k} \notin \mathrm{list}^{\mathrm{r}} \\ 0 & \text{otherwise} \end{cases}
\tag{6}
$$

where δ(N,S) is the pheromone value, Φ(N,S) is the value of the heuristic related to energy, Rs is the receiver nodes. For node N, $\mathrm{list}^{\mathrm{r}}$ is the list of identities of received data packages previously. μ and τ are two parameters that control the relative weight of the pheromone trail and heuristic value. Pheromone trails are connected to arcs. Each arc(N,S) has a trail value δ (N,S) □[0,1] . Since the destination D is a stable sink node, the last node of the path is the same for each ant travel. The heuristic value of the node N is expressed by Eq. (7):

$$\varphi(\mathrm{N, S}) = \frac{(\mathrm{E} - \mathrm{e_N})^{-1}}{\sum\limits_{\mathrm{p}\in\mathrm{Rs}} (\mathrm{E} - \mathrm{e_p})^{-1}} \tag{7}$$

where E is the initial energy, and eN is the current energy level of receiver node N. This enables decision making according to neighbor nodes' energy levels, meaning that if a node has a lower energy source then it has lower probability to be chosen. Nodes inform their neighbors about their energy levels when they sense any change in their energy levels. A special memory named Mk is held in the memory of an ant to retain the places visited by that ant (which represent nodes in WSNs). In Eq. (6), the identities of ants (as sequence numbers) that visited the node previously, are kept in the node's memories, instead of keeping node identities in ant's memories, so there is no need to carry Mk lists in packets during transmission. This approach decreases the size of the data to be transmitted and saves energy. In Eq. (6) each receiver node decides whether to accept the upcoming packet of ant k or not, by checking its list. So, the receiver node N has a choice about completing the receiving process by listening and buffering the entire packet. If the receiver node has received the packet earlier, it informs the transmitter node by issuing an ignore message, and switches itself to idle mode until a new packet arrives. After all ants have completed their tour, each ant k deposits a quantity of pheromone Δ δk(t) given in Eq. (8), where $\mathrm{J^k}$w(t) is the length of tour wk (t) , which is done by ant k at iteration t. In WSNs, $\mathrm{J^k}$w(t) represents the total number of nodes visited by ant k of tour w at iteration t:

$$\Delta \overset{k}{\delta}(t) = 1/J_w^k(t) \tag{8}$$

Pheromone values are stored in a node's memory. Each node has information about the amount of pheromone on the paths to their neighbor nodes. After each tour, an amount of pheromone trail Δ δk(t) is added to the path visited by ant k. This amount is the same for each arc(N,S) visited on this path. This task is performed by sending ant k back to its source node from the base along the same path, while transferring an acknowledgement signal for the associated data package. Increasing pheromone amounts on the paths according to lengths of tours, Jw (t) , would continuously cause an increasing positive feedback. In order to control the operation, a negative feedback, the operation of pheromone evaporation after the tour is also accomplished in Eq. (9).

$$\delta\,\mathrm{ij(t)} \leftarrow \quad \rho)\,\delta\ \mathrm{ij(t)} \tag{9}$$

In simulations, ACO parameter settings are set to values 1 for μ, 5 for τ , and 0.5 for ρ.

Table 1 Parameters for residual energy experiment

Parameter	Value
Node initial energy (□)	1 J
Length of unit data (L)	400 Bits
Unit time	60 s
Radius of network	10 m

5 Performance Evaluation

5.1 Simulation Environment

We take 9 concentrated coronas and no. of nodes in each corons are exponentially distributed toward the sink. Radius of first corona is 1 m and radius of second corona is 2 m, and third is 8 m and so on. The initial energy of each sensor is 1 J. The data generating rate of each sensor is 4×10^2 bits/min. All the parameters are listed in Table 1.

5.2 Comparison with Other Distributions

We compare our approach to Geometric distribution and uniform distribution. Geometric distribution means a non-uniform node distribution strategy to achieve nearly balanced energy depletion in the network, i.e. the number of nodes in coronas increases from corona CR - 1 to corona C1 in geometric progression with common ratio $q > 1$. In uniform distribution, nodes are distributed uniformly in the environment. Figure 1 shows the network lifetime with other distributions. We can see that the network lifetime with uniform distribution decreases with the growth of network radius. This is because the data traffic is increasing while the radius is increasing, especially for the inner coronas. Geometric distribution works better than uniform distribution. Note that the exponential distribution with ant colony optimization routing performs even better than that of geometric distribution.

Average residual energy ratios, which is the ratio of energy remained when the network lifetime ends to the sum of initial energy of all the nodes, with the other node distribution are shown in Fig. 2. We note that while the network radius is increasing the lifetime of network is decreasing, but the total initial energy is increasing, so the residual energy ratio is slowly increasing. We observe that the residual energy ratio of the network with exponential distribution with ant colony optimization routing is better than that of network with uniform distribution and that of network with geometric distribution. This implies the effectiveness of our approach.

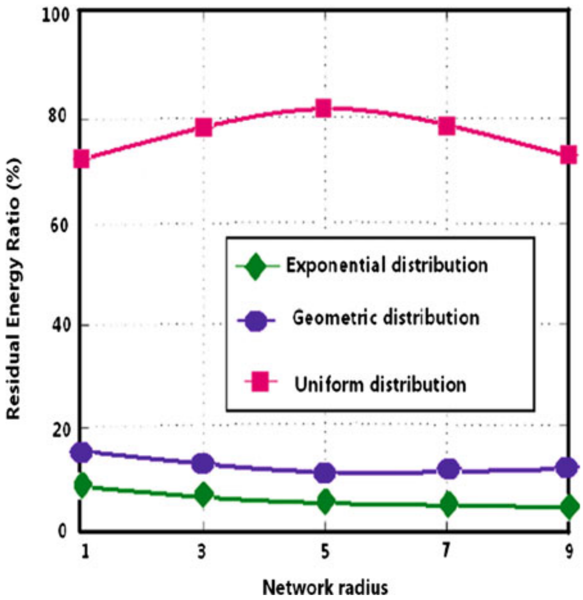


Fig. 2 Residual energy ratios of different node distributions

6 Conclusion and Future Work

In this paper we propose jointly effect of exponential distribution of nodes toward the sink and optimal transmission range so that the lifetime of wireless sensor network will be improved. In all simulations, we can see the network lifetime is significantly extended than uniform distribution and geometric distribution. A perfect MAC layer is needed to handle channel problems among the nodes. We also have to rely much on more sophisticated sensor manufacturing and sensor node deployment methods to guarantee non uniform node distribution strategy possible. Here we ignored the energy consumption on MAC and network layers which we plan to include in our future work.

References

1. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) Wireless sensor networks: a survey. IEEE Trans Comput Netw 38(4):393–422
2. AL-Karaki JN, Kamal AE (2004) Routing techniques in wireless sensor networks: a survey. IEEE Trans Wirel Commun11(6):6–28
3. Olariu S, Stojmenovic I (2006) Design guidelines for maximizing lifetime and avoiding energy holes in sensor networks with uniform distribution and uniform reporting. In: Proceedings of IEEE INFOCOM, 1–12, 2006

4. Li J, Mohapatra P (2005) An analytical model for the energy hole problem in many-to-one sensor networks. In: Proceedings of 62^{nd} IEEE Vehicular Technology Conference, vol 4. 2721–2725, 2005
5. Li J, Mohapatra P (2007) Analytical modeling and mitigation techniques for the energy hole problems in sensor networks. Pervasive Mob Comput 3(8):233–254
6. Wu X, Chen G, Das SK (2006) On the energy hole problem of non uniform node distribution in wireless sensor networks. In: Proceedings of Third IEEE International Conference Mobile ad-hoc and sensor systems (MASS '06), pp 180–187, Oct 2006
7. Lian J, Naik K, Agnew G (2006) Data capacity improvement of wireless sensor networks using non-uniform sensor distribution. Int J Distrib Sens Netw 2(2):121–145
8. Liu Y, Ngan H, Ni LM (2006) Power-aware node deployment in wireless sensor networks. In: Proceedings of IEEE International Conference sensor networks, ubiquitous, and trustworthy computing (SUTC '06), pp 128–135, June 2006
9. Shiue HY, Yu GJ, Sheu JP (2005) Energy hole healing protocol for surveillance sensor networks. In: Workshop on WASN 2005, Taoyuan, Taiwan

Multimedia Traffic Over MANETs: Evaluation and Performance Analysis

Sunil and Naveen

Abstract Providing required QoS guarantees in wireless mobile ad-hoc networks (MANETs) is much more challenging than in wired networks. This is mainly due to its characteristics such as frequently changing topology, un-centralized control, interference, multi-hop routing and collisions due to contention for channel access. Efficient and reliable routing in such an environment is very difficult task. In particular, it is essential for wireless routing protocols to provide some sort of QoS guarantees by incorporating metrics like achievable throughput, delay, jitter, packet delivery ratio, etc. In this paper, we compare and analyze the performance of various wireless routing protocols over MANETs. The source nodes are modeled using video trace files generated from real time video streams using H.264/SVC codec. Our basic goal of this paper is to analyze and exploit the in-built support for real time multimedia applications over MANETs. Performance evaluation and analysis of various wireless routing protocols over huge number of different scenarios is conducted. The results that evaluate the network performance are obtained using simulation in terms of packet delivery ratio, routing overhead and average end-to-end delay. The routing protocols are analyzed against the effects of change in network scalability (both in terms of size and number of nodes) and network load. Results obtained from simulations shows that it is possible to transmit multimedia applications traffic with limited support for video quality over MANETs.

Keywords Mobile ad-hoc routing · Performance analysis · Routing protocols · Simulation · AODV protocol · Video transmission

Sunil (✉) · Naveen
Department of Computer Science and Engineering, Suresh Gyanvihar University,
Jaipur, India
e-mail: sunil.jangir@gmail.com

Naveen (✉)
e-mail: naven_h@yahoo.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_70,

1 Introduction

Presently, there is rapid movement in infrastructure-less wireless mobile ad-hoc networks (MANET) [6] due to their mobility and inexpensive nature. With this strong need, the major focus of researchers moves towards Quality-of-Service (QoS) [6] which is must in these networks. Multimedia and other delay or error sensitive applications require a good QoS. In MANET, we have collection of mobile nodes that can communicate with each other through wireless channels. The major benefit of using MANET is that here nodes are self dependent in terms of configuring and organizing themselves according to any change in network topology. In MANETs, communication takes place directly through radio channels if communicating nodes are in the range. Each node has its own range of transmission, in case nodes are out of the range then intermediate nodes play a major role in forwarding the message.

In this paper, we perform an analysis of different routing protocols under various MANET scenarios to evaluate their performance during the transmission of multimedia traffic. Multimedia data has characteristics such as variable frame size and small inter-packet time which makes its transmission difficult as compared to best-effort traffic. We use three MANET routing protocols for our performance analysis. The routing protocols selected are as follow: (a) Ad-hoc on-demand distance vector routing protocol (AODV) [8] (a reactive routing protocol), (b) Optimized link state routing protocol (OLSR) [3] (a proactive routing protocol) and (c) Zone based routing protocol (ZRP) [7] (a hybrid routing protocol). Furthermore, We discover and discuss the issues and challenges that are faced in wireless network's during the transmission of multimedia traffic.

The remainder of the paper is organized as follows. In Sect. 2, we present an overview of related work done on the performance analysis of MANETs routing protocols. Section 3, we identify the challenges imposed by wireless networks while realizing multimedia traffic over MANETs. In Sect. 4, result evaluation through simulations and analysis is discussed. Finally concluding remarks are given in Sect. 5.

2 Related Work

In this section, we discuss the previous work done on performance analysis of routing protocol in MANETs. In [4], simulation based experiments are performed to analyze the performance of Hybrid Routing Protocols ZRP, CBRP on the basis of Packet Delivery Ratio, End to End delay and Average Throughput. These results are compared with AODV, DSR and FSR routing protocols by varying number of nodes. Although, effects of network load and network mobility are not analyzed. Also, the traffic is generated using best effort traffic applications.

Authors in [2] analyze the reactive and proactive routing protocols with varying network conditions and speed to find an optimized route from a source to some possible destination. This paper presents how routing protocol will behave in less and more stressful condition, performance of mobile ad-hoc network routing protocol such as AODV, DSDV, DSR, to simulate the above said protocol on the base of normalize routing load, throughput, Average End-to-End to delay, packet loss and packet delivery fraction.

Authors in [10] presented a detailed survey to analyze the effect of random based mobility models on the performance of Proactive Routing Protocol (DSDV—Destination Sequence Distance Vector) and Reactive Routing Protocol (AODV—on Demand Distance Vector, DSR—Dynamic Source Routing). Performance analysis is done with respect to end-to-end delay, throughput and Packet delivery ratio for varying node densities. Some other papers that have done the performance analysis of various wireless routing protocols are [11, 5].

In this paper, we have done the performance analysis of various routing protocols to know their behaviour under multimedia traffic transmission. We use trace files generated from real time video file to model the network sources in network. This way we can stress the network with real-time traffic instead of relaying on applications such as constant bit rate or variable bit rate which generates traffic in some fixed patterns. The results presented in this paper helps the researchers working in the area towards provisioning the QoS support for real-time multimedia applications. This will help to select the routing approach that can be used to create new QoS-aware routing protocols.

3 Issues and Design Challenges for Multimedia Transmission in MANETs

There are several design issues that greatly affect the performance of multimedia traffic mainly video streaming. Earlier, routing protocols in MANET does not focus on QoS for any route from source to destination. But due to rapid attraction towards MANET and evolving a vast variety of multimedia applications and many QoS-aware applications there is a strong need to involve QoS in MANET for efficient transmission of such application's traffic.

3.1 Challenges in Implementing QoS for Multimedia Traffic over MANETs

Following are the challenges we face during the provisioning of QoS-aware solutions in MANETs:

- **Irregular physical behavior**: In ad-hoc networks, we have unpredictable atmospheric and operating conditions that includes interference, thermal noise, multipath fading and shadowing. So this leads to unpredictable physical characteristics.

- **Infrastructure less and distributed architecture**: In infrastructure-less networks, each node has to send its QoS state information to other nodes, this increases complexity.
- **Communication via multi-hop**: As the nodes on an intermediate route of a source-destination pair increases the probability of route failure also increases. In a multi-hop communication, each node is important as it plays an equivalent role in forwarding the message. So, failure of any node results in communication breakdown.
- **Contention in channels**: Sharing of channel results in collision and increase in collision results in increased delay, decreased packet delivery ratio, low utilization of channel bandwidth and drain in nodes battery power. Using TDMA technique, we can remove this problem but we cannot use TDMA or CDMA in MANETs due to its de-centralized nature.

3.2 Trade-Offs in Designing QoS-Aware Solutions

The general design trade-offs that may affect the design of QoS-aware routing protocols is as follows:

- **Reactive versus Proactive Routing algorithms**: Proactive(table-driven) routing protocols aim to maintain up to-date routing information between each pair of nodes in the network and keep them consistent by transmitting routing updates at fixed time intervals. Proactive protocols produce low delay during route discovery and route set-up but consume more channel capacity.
- **Channel Capacity versus Delay**: The delay between packets can be reduced by sending more data packets through different neighbors and thus increase network capacity. If repeated packets are sent via different paths, packets are received at destination with low delay. This scheme increases network capacity on the cost of delay.
- **Transmission power: Low versus High**: Reducing transmission power increases number of hops and save energy but at other side increasing power leads to less number of hops that decreases route failure, low routing and route maintenance overhead and increases path efficiency and end-to-end reliability.

4 Simulations and Performance Evaluation

All the simulations are performed using a well known network simulator called Qualnet v5.0. The source-destination pairs selected for data transmission and reception are random. To model the source nodes with real-time video traffic, we use the video trace files [12] generated using H.264/SVC [9] codec. Furthermore, the results are generated by taking average of ten trials performed using different seed values.

In all the network scenarios, we introduce background traffic using one Constant Bit Rate (CBR) data session. The traffic generated by CBR sessions has the following properties: (a) 20 packets per second are transmitted into the network, (b) The size of each packet is 512 Bytes. Other than this, the network is configured using the simulation parameters given in Table 1.

4.1 *Effect of Increase in Network Load*

In this section, we evaluate the performance of AODV, OLSR and ZRP against increase in number of video streaming sessions in network. The number of video streaming source-destination pairs are increased by one in each scenario. In Fig. 1, we present the results for routing overhead (RO) incurred by each routing protocol for different number of video sessions in network. As we can observe from Fig. 1,

Table 1 Simulation parameters

Parameters	Values
Simulator	Qualnet 5.0.1
Simulation time	500 s
Scenario dimension	600×600, 1000×1000, 1400×1400, 1800×1800 m^2
Number of nodes	Between 25 and 100
Application layer protocol	CBR, Traffic trace
Video	Big buck bunny CIF(352×288)
Video codification	H.264 SVC
Frame size	21–3930 Bytes
Inter-packet Time	33 ms
Routing protocols	AODV, OLSR, ZRP
Mobility model	Random way-point
Data rate	11 mbps

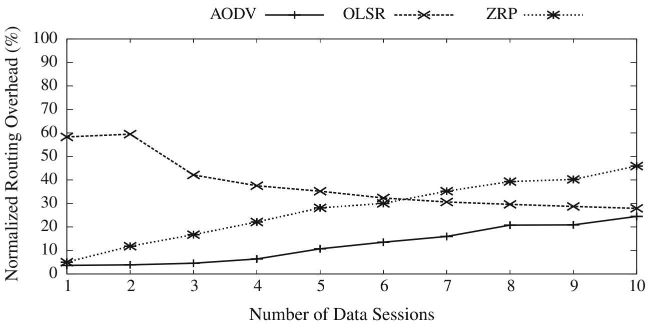


Fig. 1 Effect of increased number of data sessions on routing overhead

when number of video sessions are less in the network the RO of AODV protocol is low as compared to other routing protocols. This is because AODV maintains information about active routes only. As the number of video session increases in the network the routing overhead of OLSR decreases because it does not have to increase its RO with increase in number of video sessions. This is due to the fact that OLSR store routes for all destinations at the network start-up using its proactive routing approach, so it is not effected by any increase in source-destination pairs. On the other hand, ZRP gives intermediate RO values because it uses both, proactive and reactive route discovery processes for updating its routing tables.

Figures 2 and 3 show the effect on average end-to-end delay (EED) and packet delivery ratio (PDR) with increase in number of video sessions on AODV, OLSR and ZRP. As seen from Fig. 2 EED of AODV protocol is very low as compared to OLSR and ZRP. This is because the re-routing time of AODV is very low as compared to other routing protocols [1]. In addition to that, when we see the simulation log file for AODV, we observe that when the network traffic is increased due to increase in video sources AODV is able to find routes that are less congested as compare to other routes. Because of the above two reasons AODV protocol EED is low and PDR is high as compared to other routing protocols. On the other hand OLSR performs

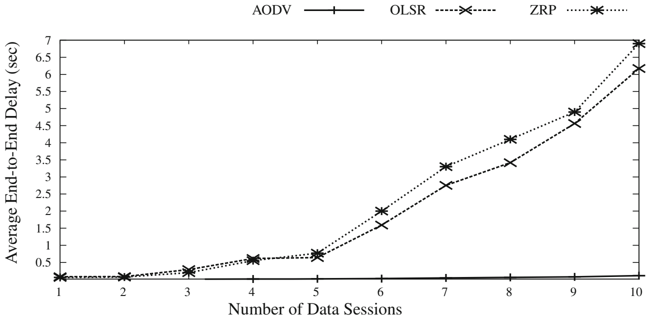


Fig. 2 Effect of increased number of data sessions on average delay

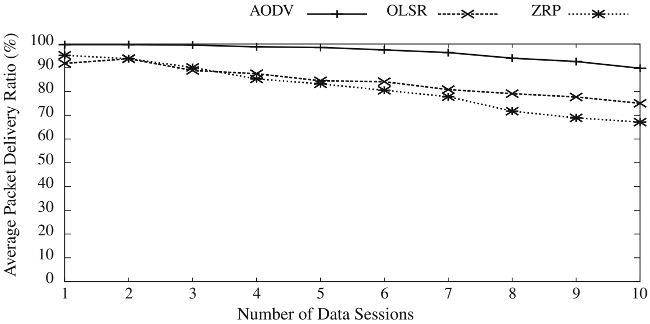


Fig. 3 Effect of increased number of data sessions on PDR

better than ZRP due to its fully proactive nature of making and updating routing table. This is because, in OLSR initial delay for route setup is very low due to the high probability of route availability for any destination is source routing table.

4.2 *Effect of Increase in Network Size*

In this section, we present simulation results obtained with increasing network scenario size in terms of number of nodes and scenario dimensions. The network is configured using the simulation parameters shown in Table 1. The number of source destination pairs in each scenario is kept constant (which is four). Figure 4 shows the RO caused by AODV, OLSR and ZRP on different sized network scenarios. As seen from Fig. 4, the RO of AODV is small as compared to OLSR and ZRP. This is because AODV only keep track of active routes so as long as the number of source-destination pairs are not changed there will be no significant change in AODV routing overhead. In addition to this, the small constant increase in AODV RO with network size is due to the increase in number of intermediate nodes in an active route. On the other hand, RO of OLSR increases greatly with increase in network size because of its proactive route discovery process due to which the number of control message broadcast increases greatly. ZRP performs slightly better than OLSR due to its semi-reactive route discovery process.

In Figs. 5 and 6, we present EED and PDR results with change in network size for AODV, OLSR and ZRP routing protocols. As we can observe from Fig. 5, the EED of AODV is less as compared to OLSR and ZRP. This is due to two reasons: (a) the size of routing table is low in AODV due to its reactive route discovery process. This helps to get a route for a destination from routing table very quickly as compared to routing tables that are build using proactive route discovery methods due to their large sizes. (b) As the size of network increases the intermediate routes between a source destination also increases which further increases the probability

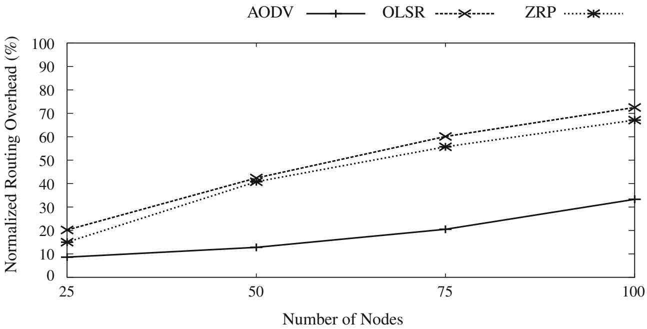


Fig. 4 Effect of increase in network scalability on routing overhead

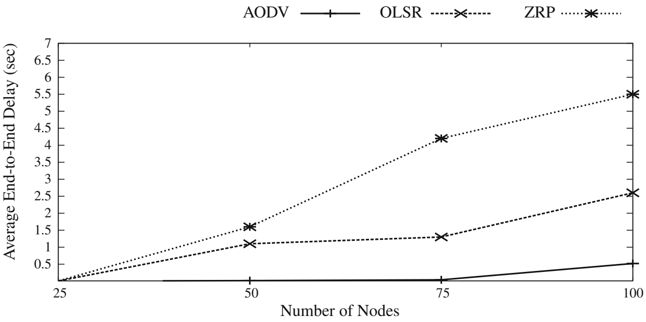


Fig. 5 Effect of increase in network scalability on average delay

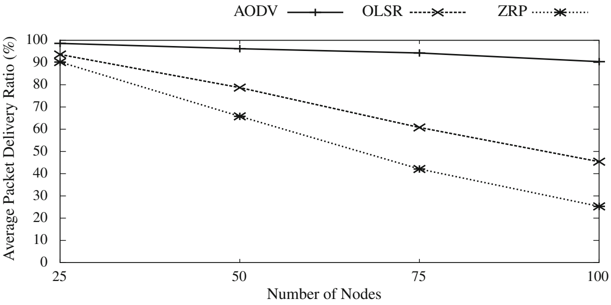


Fig. 6 Effect of increase in network scalability on PDR

of link failure. So, as mentioned above that the AODV has low re-routing time it will recover from these frequent route breaks quickly as compared to OLSR and ZRP. This will greatly minimizes AODV EED. In Fig. 6, we can observe that as the network size increases PDR starts decreasing due to increase in routing overhead and increase in route breaks caused by high number of intermediate nodes.

5 Conclusion and Future Work

In this paper, we have done performance analysis of various MANET routing protocols to evaluate their behaviour during transmission of multimedia applications. We have present and analyze results for different routing protocols in different scenarios with varying network load and network size. We observe from simulation results that reactive routing protocols (such as AODV) are well suited for MANET like environments due to their low routing overhead, low re-routing delay and adaptability towards dynamically changing network conditions due to node mobility.

References

1. Calafate CT, Malumbres MP, Manzoni P (2008) Performance of H.264 compressed video streams over 802.11b based MANETs. In: Proceedings of the 24th international conference on distributed computing systems workshops, 2004, March 2008, pp 776–7812
2. Chaudhary MS, Singh V (2012) Simulation and analysis of routing protocol under CBR and TCP traffic source. In: Communication systems and network technologies (CSNT), 2012 International Conference, May 2012, pp 342–346
3. http://tools.ietf.org/html/draft-cole-manet-olsrv2-mib-00
4. Khatkar A, Singh Y (2012) Performance evaluation of hybrid routing protocols in mobile ad hoc networks. In: Advanced computing communication technologies (ACCT), 2012 second international conference, Jan 2012
5. Kuppusamy P, Thirunavukkarasu K, Kalaavathi B (2011) A study and comparison of OLSR, AODV and TORA routing protocols in ad hoc networks. In: Electronics computer technology (ICECT), 2011 3rd international conference, April 2011, pp 143–147
6. Morten L, Kristiansen S, Plagemann T, Goebel V (2011) Challenges and techniques for video streaming over mobile ad hoc networks. Multimedia Syst Springer 17(1):51–82
7. Pearlman MR, Haas ZJ (1999) Determining the optimal configuration for the zone routing protocol. Sel Areas Commun IEEE J 17(8):1395–1414
8. Perkins C, Royer E (1999) Ad hoc on-demand distance vector routing. In: 2nd IEEE workshop mobile computation system applications, 1999
9. Rimac-Drlje S, Nemcic O, Vranjes M (2008) Scalable video coding extension of the H.264/AVC standard. ELMAR (2008) 50th international symposium. Sept 2008, pp 9–12
10. Rohankar R, Bhatia R, Shrivastava V, Sharma DK (2012) Performance analysis of various routing protocols (proactive and reactive) for random mobility models of Adhoc networks. In: Recent advances in information technology (RAIT), March 2012, pp 331–335
11. Tan Y, Song Y, Zhou S (2011) Simulation based comparative performance analysis of on-demand routing protocols. In: Smart and sustainable city (ICSSC 2011), IET international conference, July 2011, pp 1–5
12. Video Trace:http://trace.eas.asu.edu/

Web Accessibility: Designing and Testing of Web Based Application for Persons with Disabilities

Kalpana Johari and Arvinder Kaur

Abstract The latest technological development in web and mobile communication technology has made access to the Internet on the Internet enabled mobile handset quite simpler. But this rapid technological growth has not touched all the sections of the society especially the under privileged ones like the economically weaker strata of society and the people with disabilities. Therefore, in this paper we make an effort to design and develop the Web based application for the Physically Challenged Persons, on the mobile handsets. Our primary focus is to understand the needs and requirements of the Physically Challenged Persons and then design the application which they can efficiently use on the mobile handsets.

1 Introduction

The access of web contents and web application from Internet on mobile devices has become a rule rather than exception as more and more people are accessing Internet on their mobile handsets such as smart phones, iPhone, tablet, palmtop etc. The web contents are primarily designed to be accessed on desktop based browser application. Although the fundamental idea behind Internet was accessibility of web contents to all the users having or not having any disability, accessing it on any type of Internet enabled hardware or software systems. The mobile devices being constrained and variant in terms resources such as memory, computing power, browsers, screen size, input mechanism, way of accessing Internet etc becomes more susceptible

K. Johari (✉)
Centre for Development of Advanced Computing, Noida, India
e-mail: kalpanajohari@cdoc.in

A. Kaur (✉)
USICT, GGSIP University, Dwarka, Delhi 110078, India
e-mail: arvinder70@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_71,

to accessibility related issues. Therefore, accessibility of web contents on mobile devices has picked up much attention of research community.

It is important for the web designers and developers to measure and evaluate the accessibility of the web contents and web application before being published. The approaches used to evaluate the accessibility of web contents on desktop based browser cannot be applied for evaluating the accessibility of web contents on mobile devices. This is mainly due to the fact that such methodologies are meant to evaluate the accessibility from the point of view of people having some form of disability and they do not take into consideration the limitation of the devices. Moreover, such approaches are designed around the Web Contents.

Accessibility Guidelines (WCAG1.0 and WCAG2.0) [1, 2] designed by W3C. The methodologies used for evaluating the accessibility of web contents are either based on usage of automated tools [3–5] or use metrics specially designed for evaluating accessibility [6–10]. These automated tools although provide scalability, they fail firstly, to fully validate the conformance of web contents, with respect to guidelines and secondly, to capture the actual scenario of human accessing the web contents [11].

In [12] Harper suggested that the accessibility issues that people, with some form of disability, encounter are similar to the one that people while accessing web contents on mobile devices face. Therefore the techniques meant to evaluate web accessibility with respect to people having some disability can also be applied to evaluation of mobile web accessibility. We do not totally agree with this thought because in case of mobile devices making the contents available on resource constrained devices itself is a challenge where as in case of making the web accessible to people with disability, the challenge lies in presenting the available contents in such a manner so that it can be accessed by people with disability. In case of mobile devices because of resource constraints and the process involved in conversion of HTML web contents to WML contents and finally delivering the contents to mobile devices has to be handled which is not the case when dealing with web accessibility issues pertaining to people with disability. For example in order to make web contents accessible to people having cognitive disability it is important to present it in a manner so that it can be perceived by them. While dealing with mobile devices a web page containing scripts, tables or lot many images of various formats may become in accessible due to lack of support from the mobile device itself. For mobile devices inaccessibility mean non availability of contents where as for people with some disability inaccessibility means contents not perceived. Therefore we believe that we require approaches, other than the one meant for evaluating web accessibility, to evaluate mobile web accessibility.

One of the earliest attempt to evaluate the accessibility of Internet based web contents on Mobile devices is made by Best Practices Working Group (BPWG) in the form of Mobile Web Best Practices (MWBP) guidelines [13] proposed as a result of Mobile Web Initiative (MWI). Following these guidelines number of attempts have been made to evaluate the mobile web accessibility. Most of the approaches meant to evaluate mobile web accessibility rely on automated tool that verify the accessibility on the basis of MWBP guidelines [14–17]. These tools suffer from the same problems as those developed to evaluate web accessibility.

2 Related Work

One of the earliest attempt to evaluate the mobile web accessibility is made by Arrue et al. [15]. They have presented a tool based approach to evaluate accessibility of mobile web. They have extended a tool EvalAccess, meant to evaluate web accessibility as per the WCAG1.0, to EvalAccess Mobile on the basics of MWBP [13]. The tool is available as a web service. Since all the guidelines presented in WCAG1.0 and MWBP cannot be automated therefore, the tool only provides check for very limited guidelines. Moreover the tool does not provides the report on the accessibility of the complete user scenario and therefore the effectiveness of the tool at evaluating the accessibility of mobile web contents cannot be established. Similarly, the other tools based approach proposed in [14, 16] suffer from the same issues.

In [18] Roto et al. have identified and presented the major features that affect the user's browsing experience on mobile devices. According to the authors it is not just the web contents but various other factors such as network infrastructure, user's state, mobile device, browser etc. that determines the accessibility of web contents on mobile devices.

Hori et al. in [11] the authors have proposed the inspection based evaluation approach. They have pointed the need for scenario based evaluation of accessibility of web contents. Also the perceivable and cognitive aspect of accessibility has been highlighted by the authors. The problem with inspection based evaluation is that it is not easily scalable and therefore may be difficult to implement on large web based system.

Vigo et al. [17, 19] have highlighted the need to evaluate the accessibility of web contents by considering the device features. The approach presented by the authors is a tool based approach using the device descriptions. The approach is good in the sense that it takes into consideration the affect of characteristics of devices but the user aspect at evaluating the accessibility is missing in the approach. Moreover the technique extends the Mobile OK basic test which itself partially evaluate the MWBP1.0.

The metrics pertaining to Web application for the desktop applications has been reported by number of authors. The work of Mendes et al. [20] on web metrics mainly deals with the estimation of authoring efforts. The quality metrics reported Ben Lilburne et al. [6] presents a framework on three quality attributes, usability, reliability and maintainability of web application. The work by Fewster et al. [21] on measurement prediction and risk analysis is again relevant to web applications for desktop machines. The details of the metrics developed for web application is reported in [7]. The quantitative metrics reported in [10] is meant for information retrieval and accessibility monitoring for web application. The work on evaluation of web accessibility [8] proposes the web accessibility metrics for people with special needs. To the best of our knowledge proposal of mobile metrics to measure the accessibility of web contents is the first reported work. The related work reported in this section is not exhaustive but seemed relevant to our work. The issues in accessing the web application and the web pages on the desktop applications are quite different from the one in Mobile handsets and therefore justifies the need of formalization of the metrics.

3 Guidelines and Norms

In order to bring technology to the doorstep of the physically challenged persons so as to integrate them in a mainstream society, the W3C has proposed exhaustive set of good practices to be followed, and a set of procedures, guidelines and standards on making Web content accessible to the people either through Desktop Computer/Laptop or by harnessing the power of handheld devices such as iphones and ipads. Eversince the UN Convention on the Rights of Persons with Disabilities (2006) recognized Web accessibility as a basic human right various set of guidelines namely WCAG, MWBP, Web Accessibility Initiative (WAI) and Mobile Web for Social Development (MW4D) guidelines has been proposed and formulated. WCAG is a guide for making Web sites accessible to people with physical and mental disabilities, Mobile Web Best Practices (MWBP) is a guide for making Web sites usable from a mobile device. Therefore, the Primary objective of both WCAG and MWBP is to improve the interaction of the common men with World Wide Web and try to figure out the difficulties he faces either due to his physical or mental disabilities or by the usage of the hand held devices he uses to access the World Wide Web. It is always emphasized that the Web be accessible to everyone in order to provide equal access and equal opportunity to people with disabilities. In our work we develop software applications conforming to the guidelines and standards as suggested in WCAG, MWBP and MW4D for iphones and ipads.

4 Methodology Adopted

Given below are the sequence of steps which we adopted to study the problems faced by PWD (Persons with Disabilities) in their day to day life while using the technology aided hand held devices like iphones, ipads and other devices:-

(1) To better understand their needs and special requirements we visited two well known Centers for physically challenged persons:- AADI (Action for Ability Development and Inclusion), All India Deaf and Dumb Society (AIDDS) in Delhi.

(2) After having first hand knowledge of their problems,for our empirical study we prepared a questionnaire comprising a set of 15 questions [22] and went back to the respondents, got their answers/responses which can be viewed at (URL: http://bit.ly/Ik1O16). These responses helped us a lot to understand the challenges they face in using the mobile devices.

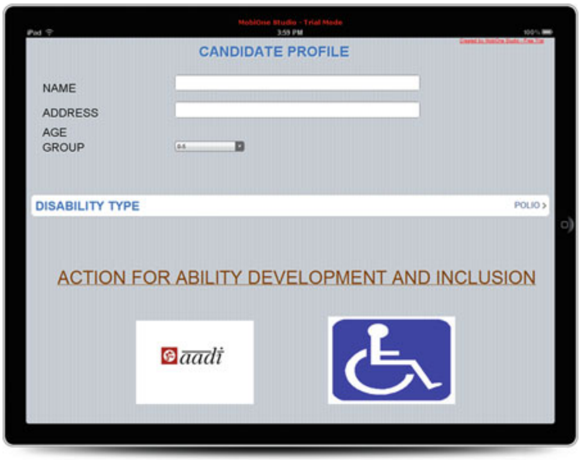
(3) Keeping their demand and needs we designed the mobile based applications using Open Source internationally acceptable toolkit called MOBIONE both for iphones and ipads. The applications thus designed are hosted on the MOBIONE Server. One can view them as described in Table 1.

(4) The Snapshots of the applications designed are thus presented in Figs. 1 and 2.

(5) In order to ensure that our applications are compliant with the international standards for Web accessibility: namely Mobile Web Best Practice WCAG2.0 we got our application tested and approved by running it multiple number of times and getting certified by W3C mobileOK Checker utility (http://validator.w3.org/mobile). The W3C mobileOK Checker follows the W3C mobileOK Basic Tests 1.0 standard and uses specific HTTP headers to retrieve resources as if it were a mobile device [As per W3C Mobile Ok Site].

(6) The results of our application are presented in Table 1.

(7) As our applications were developed using Open Source tool kit, so we distributed free of cost these applications to the centre we visited. We have given them a user manual on how to use our applications, we trained them how to use and we gave them eight (8) week feedback period where in again we would obtain their feedback, and develop next generation of the mobile application adhering to international norms and practices and meeting to the demands of the Physically challenged Persons.



CANDIDATE PROFILE

NAME

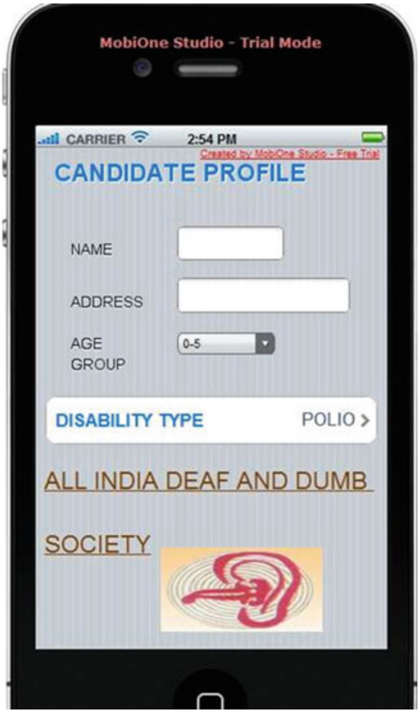
ADDRESS

AGE GROUP 0-5

DISABILITY TYPE POLIO >

ACTION FOR ABILITY DEVELOPMENT AND INCLUSION

Fig. 1 Snapshot for iPad application


MobiOne Studio - Trial Mode
CARRIER
2:54 PM
Created by MobiOne Studio - Free Trial

CANDIDATE PROFILE

NAME

ADDRESS

AGE GROUP 0-5

DISABILITY TYPE POLIO >

ALL INDIA DEAF AND DUMB SOCIETY

Fig. 2 Snapshot for iPhone application

Table 1 Results after testing an application

S. No.	Device used	Organization	Available at URL	Results by W3C mobile checker (%)
1	iPhone	AIDDS	http://validator.w3.org/mobile/check?task=2012073009270920.mobile2 docAddr=http%3A%2F%2Fgoo.gl %2F4XP5G	72
2	iPad	AADI	http://validator.w3.org/mobile/check?task=20120730103227143.mobile1 docAddr=http%3A%2F%2Fgoo.gl %2Fma60e	69

5 Conclusion and Future Work

We propose to extend this work by designing more Web Applications with other toolkits such as Nokia Mobile Internet Toolkit for Desktop and laptop Computers and NOKIA WAP Toolkit for the PWD especially keeping in mind their special needs.

In order to ensure that our applications are compliant with the international standards for Web accessibility: namely Mobile Web Best Practice and WCAG2.0 we propose to get our application(s) tested and approved by running it multiple number of times and would get it certified by W3C mobileOK Checker utility (http://validator.w3.org/mobile), so as to ensure that applications which we designed are in lines with international standards.

Acknowledgments I wish to express sincere gratitude to the administration of GGSIP University, Center for Development of Advanced Computing for providing the academic environment to pursue research activities. Our heart full thanks and sincere gratitude to the students, staff member, counselors and administration of to the AADI (Action for Ability Development and Inclusion) and All India Deaf and Dumb Society (AIDDS) for providing us fine inputs and sparing their time and energy for timely feedback and suggestions.

References

1. Caldwell B, Cooper M, Reid LG, Vanderheiden G (2008) Web content accessibility guidelines 2.0. World Wide Web Internet And Web Information Systems, 0(May 1999):1171–1172. http://www.w3.org/TR/WCAG20/
2. Chisholm W, Vanderheiden G, Jacobs I (1999) Web content accessibility guidelines 1.0. Interactions 8(4):35–54. http://portal.acm.org/citation.cfm?id=379550
3. Abascal J, Arrue M, Fajardo I, Garay N, Tomas J (2004) The use of guidelines to automatically verify web accessibility. Univers Access Inf Soc 3:71–79. doi:10.1007/s10209-003-0069-3
4. Benavídez C, Fuertes J, Gutiérrez E, Martínez L (2006) Semi-automatic evaluation of web accessibility with hera 2.0. In: Miesenberger K, Klaus J, Zagler W, Karshmer A (eds) Computers helping people with special needs. Lecture notes in computer science, vol 4061. Springer, Berlin, pp 199–206. doi:10.1007/1178871330
5. Rowan M, Gregor P, Sloan D, Booth P (2000) Evaluating web resources for disability access. In: Assets '00: proceedings of the fourth international ACM conference on assistive technologies. ACM, New York, pp 80–84. doi:10.1145/354324.354346
6. Lilburne PDB, Khan KM (2004) Measuring quality metrics for web application. Technical report no: CIT/08/2004, University of Westen Sydney
7. Calero C, Ruiz J, Piattini M (2004) A web metrics survey using wqm. In: Koch N, Fraternali P, Wirsing W (eds) Proceedings of the web engineering—4th international conference (ICWE 2004) Munich, Germany, 26–30 July 2004. Lecture notes in computer science, vol 3140. Springer, Heidelberg, pp 147–160
8. Freire AP, Fortes RPM, Turine MAS, Paiva DMB (2008) An evaluation of web accessibility metrics based on their attributes. In: Proceedings of the 26th annual ACM international conference on design of communication SIGDOC '08. ACM, New York, pp 73–80. doi:10.1145/1456536.1456551
9. Parmanto B, Zeng X (2005) Metric for web accessibility evaluation. J Am Soc Inf Sci Technol 56(13):1394–1404. doi:10.1002/asi.20233
10. Vigo M, Lomuscio R, Arrue M, Abascal J, Brajnik G (2007) Quantitative metrics for measuring web accessibility. In: Proceedings of the 2007 international cross-disciplinary workshop on web accessibility (W4A). ACM Press, New York, pp 99–107
11. Hori M, Kato T (2008) Mobile web and accessibility, pp 302–313. http://www.springerlink.com/content/j3g53r7533m06k1g
12. Harper S (2008) Mobile web: reinventing the wheel? SIGACCESS Access Comput (90):16–18. doi:10.1145/1340779.1340781

13. Sullivan B, Connors A (2010) Mobile web best practices. W3C, March 2010. http://www.w3.org/TR/mwabp/
14. Tobar LM, Andrés PML, Lapena EL (2008) Weba: a tool for the assistance in design and evaluation of websites. J Univers Comput Sci 14(9):1496–1512
15. Arrue M, Vigo M, Abascal J (2007) Automatic evaluation of mobile web accessibility. In: ERCIM'06: proceedings of the 9th conference on user interfaces for all. Springer, Berlin, pp 244–260. http://dl.acm.org/citation.cfm?id=1783789.1783808
16. Bandeira R, Lopes R, Carrio L (2010) Towards mobile web accessibility evaluation. In: Proceeding of the ETAPS 2010 FOSSAMA workshop
17. Vigo M, Aizpurua A, Arrue M, Abascal J (2008) Evaluating web accessibility for specific mobile devices. ACM Press, New York, pp 65–72. http://portal.acm.org/citation.cfm?id=1368044.1368059
18. Roto V (2006) Web browsing on mobile phones—characteristics of user experience web browsing on mobile phones—characteristics of user experience doctoral dissertation. Technology 152(3):86+60. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.98.5091
19. Vigo M, Aizpurua A, Arrue M, Abascal J (2009) Automatic device-tailored evaluation of mobile web guidelines. New Rev Hypermedia Multimedia 15(3):223–244. http://www.tandfonline.com/doi/abs/10.1080/13614560903473886
20. Mendes E, Mosley N, Counsell S (2001) Web metrics: estimating design and authoring effort. IEEE MultiMedia 8(1):50–57. doi:10.1109/93.923953
21. Fewster R, Mendes E (2001) Measurement, prediction and risk analysis for web applications. In: Proceedings of the 7th international symposium on software metrics, METRICS'01
22. Johari K, Kaur A (2012) Measuring web accessibility for persons with disabilities. In: Fourth international conference on computational intelligence and communication networks (CICN 2012)

CBADE: Hybrid Approach for Duplication Detection and Elimination

A. Anny Leema, P. Sudhakar and M. Hemalatha

Abstract Radio Frequency Identification (RFID) is an automatic data capturing technology and the dirty data stream generated by the RFID readers is one of the main factor limit the widespread adoption of RFID technology. In order to provide reliable data to RFID application, it is necessary to clean the collected data before they are subjected to warehousing. In this paper we are going to construct the elegant hospital environment using RFID and developed the cellular based approach algorithm to clean the duplication anomaly. First middleware approach is applied to deal with low complex anomalies and in next stage deferred approach is followed to deal with high complex anomalies based on business context. Simulation shows our cleansing approach for duplication removal deals with RFID data more accurately and efficiently. Thus we can bring down the health care costs, optimize business processes, streamline patient identification processes and improve patient safety.

1 Introduction

RFID (radio frequency identification) technology uses radio waves to transfer data between readers and movable tagged objects. In a networked environment of RFID readers, enormous data is generated from the proliferation of RFID readers. In RFID

A. Anny Leema (✉)
Karpagam University, Coimbatore, India
e-mail: annyleema@gmail.com

P. Sudhakar
Department of Computer Science and Engineering, M. Kumarasamy College of Engineering, karur, India
e-mail: navaladiperiyasamy@gmail.com

M. Hemalatha
Department of Software Systems, Karpagam University, Coimbatore, India
e-mail: hema.bioinf@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_72,

environment, the database becomes more pervasive, therefore, various data quality issues regarding data legacy, data uniformity and data duplication arise. The raw data generated from the readers can't be directly used by the application. Thus, the RFID data repositories must cope with a number of quality issues. These data quality issues include data redundancy, false positive and false negative. Data quality has become increasingly important to many organizations [1]. This is especially true in the health care field where cost pressures and the desire to improve patient care drive efforts to integrate and clean organizational data.

The RFID technology has major benefits as follows:

- RFID technology can recognize information on multiple products at the same time as a long-sensing range.
- Mobile tracking devices can be reused or disposed, as the RFID operation requires.
- RFID does not require line-of-site communications between a receiver and a transmitter. This fact increases the range of RFID applications.
- RFID can work in the very harsh environment.
- RFID can function with low maintenance cost and without human interaction.
- Unlike barcodes, certain RFIDs can store data, allowing changes in the objects handling and processing.

2 RFID Data and its Components

Data generated from an RFID application can be seen as a stream of RFID tuples of the form (EPC; location; time) where EPC is a unique identifier code read by an RFID reader, location is the place where the RFID reader that reads the Tag and time is the reader captures the Tag's EPC. Tuples are stored based on the chronological Timestamp. An RFID reader scanning of tags can either be programmed to work at a fixed time interval or on a continuous basis.

RFID composed of three components—an interrogator (reader), passive tag(s), and a host as shown in Fig. 1. Among the types of tags—passive, active and semi passive, passive tags have much demand due to their least system cost and long life. The tag is composed of an antenna coil and a silicon chip that includes basic modulation

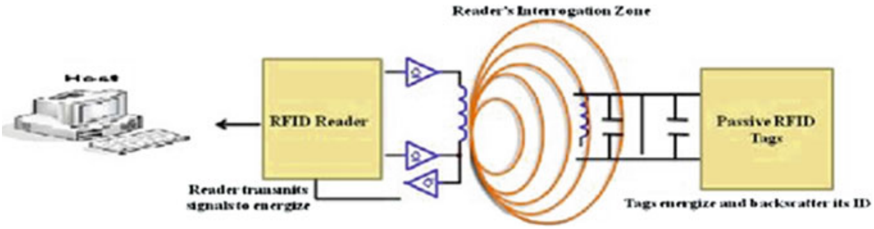

Reader's Interrogation Zone
Host
RFID Reader
Passive RFID Tags
Reader transmits signals to energize
Tags energize and backscatter its ID

Fig. 1 RFID and its components

circuitry and non-volatile memory. The tag is energized by a time-varying electromagnetic radio frequency (RF) wave that is transmitted by the reader. When the RF field passes through an antenna coil, an AC voltage is generated across the coil which is rectified to supply power to the tag. The tag using the mechanism of backscattering transmits its ID to the reader. By detecting the backscattering signal, the reader demodulates the received signal to retrieve tag's ID.

3 RFID System Design

RFID System design is depicted in Fig. 2. Raw RFID data stream is a spontaneous and very complex data to use for any analysis. The Readers are the detection nodes and are deployed in different locations across various places in the Hospital. Each detection node is identified by a unique ID that serves as the location ID. RFID tags in different locations are detected by these readers. One of the biggest challenges of the RFID data is the data volume. Sending terabyte data in to a centralized system for data cleaning requires a high performance server as well as a high speed network, which will inevitably increase the total hardware cost. Some of the data cleaning methodologies apply to data fetched by the readers, some requires an RFID middleware and others require a centralized data processing server to handle the raw data. The server level data observations include data validations, data inconsistencies and identification of anomalies before entering the enterprise application database. Data Inaccuracies are inevitable in the RFID system considering the complexity of deployment and diverse business needs it caters to.

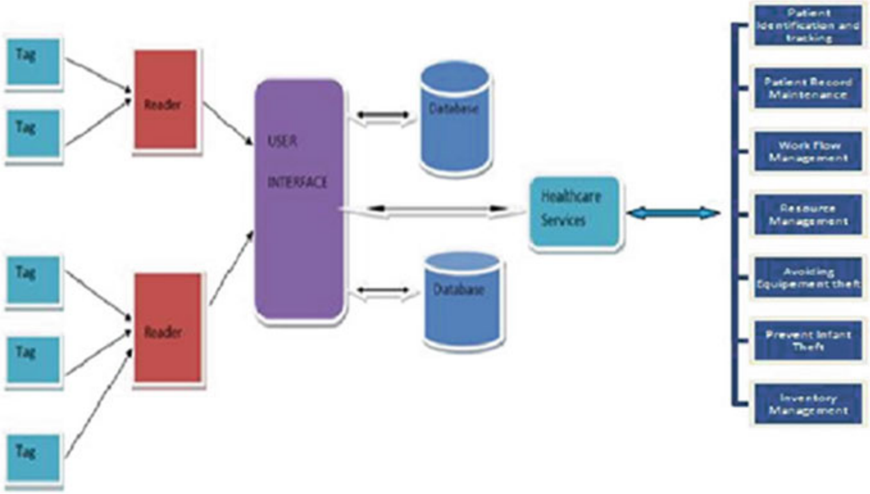


Fig. 2 RFID system design

4 Smart Hospital Environment

This paper discusses how Radio Frequency Identification (RFID) technologies were used in a hospital to construct smart environment [2, 3]. The accuracy and cost effectiveness of RFID is not guaranteed because of the issues we discovered by deploying this technology. Another issue involved managing the volume of data collected during the day. It is decided to record all the values associated with each tag event for future reference otherwise too much valid data will be lost. Finally the management can analyse the data and filter by applying business rules based on the requirement.

5 Issues in RFID Data

In the working mechanism of RFID reader and RFID tag, raw RFID readings are not always reliable [4]. Other causes of failed reads include the presence of metal in the tag vicinity, since it distorts the magnetic flux, thus weakening the energy coupling to the tag [5].

5.1 False Positive Error

An RFID reader periodically sends out RF signals to its range. When an RF tag that moves within the range of the reader receives the signals, it will send response signal along with its unique identifier code, timestamp and location ID. The reader receives the response signal and will register the data stream as one entry.

There would be some RF tags which is not supposed to be detected by the reader may be read due to the spatial divergence of RF signals sent by the reader. Such readings are termed as false positive readings.

These are predominantly caused by the RFID tags outside the normal reading scope of a reader, captured by the reader or unknown reasons from the reader or environment. The information is stored periodically by the reader through the middleware application to the database.

5.2 False Negative Error (Missing Data)

The raw RFID data streams do not provide a correct representation of the physical world that they are representing [6]. A significant number of tags which are within the reader's read range are not consistently read by the reader due to either their

orientation with respect to the reader, distance from the reader, presence of metal, dielectric or water material close to the tag and other factors [7].
Although theoretically speaking, all the tags would be read seamlessly and simultaneously on every read cycle; practically few of the tags might not be read in every cycle though present in the effective detection range. Such missed readings are deemed as false negative readings or missed readings.
These problems are very common in RFID applications and often happen in a situation of low-cost and low-power hardware [8], and the detection ability of a reader and environmental constraints. These missing tags imply that typically only a considerable percent of the tag population is actually observed on every read cycle.

5.3 Redundancy in RFID System

(i) Reader Level Redundancy

Redundancy at the reader level is the result of Cellular architecture where overlapping of reader coverage range is unavoidable. This problem occurs when a tag is in the vicinity of more than one reader at a specific time. As all the tags communicate simultaneously with the readers by sending RF signals, two are more reader reads the data from a single Tag.
Consider a scenario where readers R1, R2 and R3 are redundant since the tag T1 is read by all three readers at the same time thus responsible for reader level redundancy.

(ii) Data Level Redundancy

RFID data are usually streams of data and hence the redundancy on the data level has always been handled in the general way of dealing with data streams which might not be the optimal solution considering the uniqueness of RFID data. RFID applications handle humongous data as some readers can read hundreds of tag readings in a second. RFID data stream is considered as spontaneous and periodical in nature. The RFID data on average is highly useful than other data streams. The less useful part of RFID data is the data that are continuously reported after the initial reading.

For instance, in Hospital Management System, a tagged entity, (Say a doctor) may move to his consulting room and sit the whole day and send the data to the RFID management system constantly through the reader placed in his vicinity. But, from the management point of view, the most useful information for event detection is when the tagged entity (Say a Doctor) enters and exits his consulting room. Therefore, it is necessary to reduce RFID data redundancy before processing.

6 RFID Middleware

It serves as an interface between the RFID reader and Hospital management system database. RFID middleware serves as a platform which performs intelligent grouping of raw RFID data under predefined categories, to an extent filter raw RFID data stream based on anomalies, redundancy and preconditions [9, 10]. It is also responsible for mapping the low-level data stream from readers to a more manageable form that is suitable for application level interactions. The modules responsible for this mapping were called Savants in the original EPC work. Savants may be likened to the wrappers used in data integration systems and “edge systems” [11].

RFID middleware layer empowers healthcare providers by providing valuable data with a prompt connectivity. Application-level filtering based on exclusive process followed in each of the hospital services, validating data at different levels to ensure data consistency, monitor incoming data stream, provide real-time integration with the existing hospital management system, mapping data on to the relevant database table, redefining and executing business rule set are the various prime functions of the RFID middleware system. Figure 3 depicts this middleware concept. Data loading and extraction is the key to data intensive systems like RFID system.

6.1 *Cleaning Data at the Time of Insertion*

Step 1: Data insertion for source. (RFID Readings)
Step 2: Compare Input data stream with allowed data/character types. (Null, Alpha, Numeric, symbols)
Step 3: Check occurrence of similar incoming data streams for each set of streams using a for loop to identify data duplication.
Step 4: Display set of all duplicate data streams.
Step 5: Ability for the system to identify and delete redundant records with option for manual deletion.

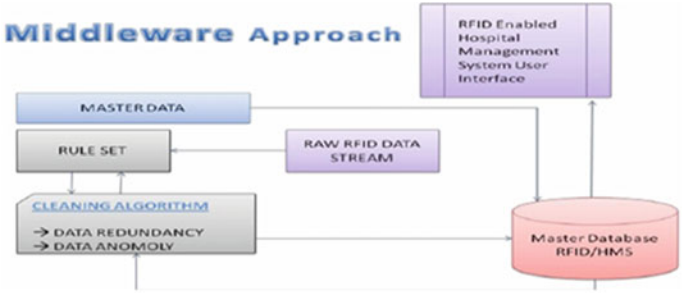


Fig. 3 Middleware approach

6.2 Identify Redundant Data Using Geometric Approach

Step 1: Get input data stream values
Step 2: Scan records in each of the rows
Step 3: Increment row index and perform search
Step 4: Automate search using for loop
The above said Geometric approach work finely to search a single value and does not suit for the huge enormous RFID data. So our approach is cellular based depicted in the Fig. 3. The diagram specifies different departments with unique Reader id which is location id.

6.3 Proposed Methodology and its Architecture

Proposed Approach is hybrid approach of middleware and deferred and the premise chosen is cellular based for detecting out of the range readings.
The RFID readers have Omni-directional antenna and hence there are possibilities for the adjacent regions to overlap with each other.
The Chosen premise to test our algorithm is depicted in the Fig. 4 and the architecture diagram of the proposed methodology is depicted in the Fig. 5.

6.4 Advantages of Proposed Approach

It is not always possible to remove all anomalies and redundancies beforehand. The rules and the business context required for cleansing may not be made available at the data loading time.

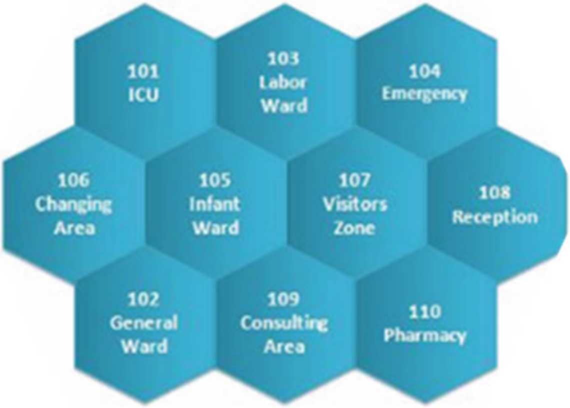


Fig. 4 Premise

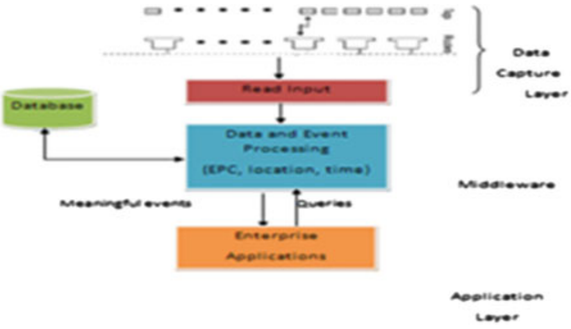


Fig. 5 System architecture

Maintaining multiple cleaned versions of RFID data are prohibitive and worthless when the rule sets defined by the application is dynamic in nature.

Known anomalies duplication are detected and corrected in the middleware approach and the meaningful data are sent to the database where processing of other anomalies is deferred until the query time.

Each application specifies the detection and the correction of relevant anomalies using declarative sequence-based rules. An application query is then automatically rewritten based on the cleansing rules that the application has specified, to provide answers over cleaned data [12].

6.5 Proposed CBADE-Cellular Based Approach Algorithm for Duplication Detection and Elimination

```
Algorithm CBADE ( Reader[ ], Tag [ ])
// Input: Reader R1 , Reader R2
// Input: Tag 1, Tag2, Tag3……Tag n
Begin
For (every tag in reader X (X=A, B))do
If
count (Selected Tag_Id) in all Tag_id > 1
Sub (Each Similar Tag_Id timestamp - Select TagId Timestamp)=0;
Return Duplication detected;
Select Max (Tag Id Timestamp);
Delete Other Tag_Ids;
else
return No Duplicates;
end for
End
```

7 Simulations

Simulation has long been used as a decision support tool in various sectors. There are several examples of simulation studies of healthcare facilities in the literature. Simulation is especially suited to the analysis of healthcare organizations due to its ability to handle high complexity and variability which is usually inherent in this sector [13]. It also acts as continuous quality improvement framework by integrating with the software agent developed via a database structure. Experimentation of different workflows, staffing decisions and what-if analysis are all promising applications of simulation in healthcare and it is practically infeasible in a healthcare environment. Simulation study requires deliberate data collection effort over a considerably long period of time. We have developed a simulator designed in C# with SQL Server 2008 as backend to evaluate the performance of the proposed algorithm. For simulation, the project premise chosen is given in Fig. 4 one RFID reader is assigned for each department with 100 tags in its interrogation zone is considered. The reader is modelled based on the design features of SkyeTek's M1—Mini RFID Reader. This reader operates from a Lithium rechargeable battery which has 0.48 KJ of energy. The tag to reader data rate is taken as 26 Kbps as per ISO 15693.
The following Assumptions to be made:

- Tag's antenna is never at 90°.
- RFID Readers are allowed to transmit energy until all tags are read.
- Communication from Tag to Reader is modeled as Poisson process.
- Reader has the knowledge on the number of bits present in a tag ID.
- Reader is unaware of the number of tags.
- Although tags are energized at the same time, the energy consumption is estimated only after the reading process has started.

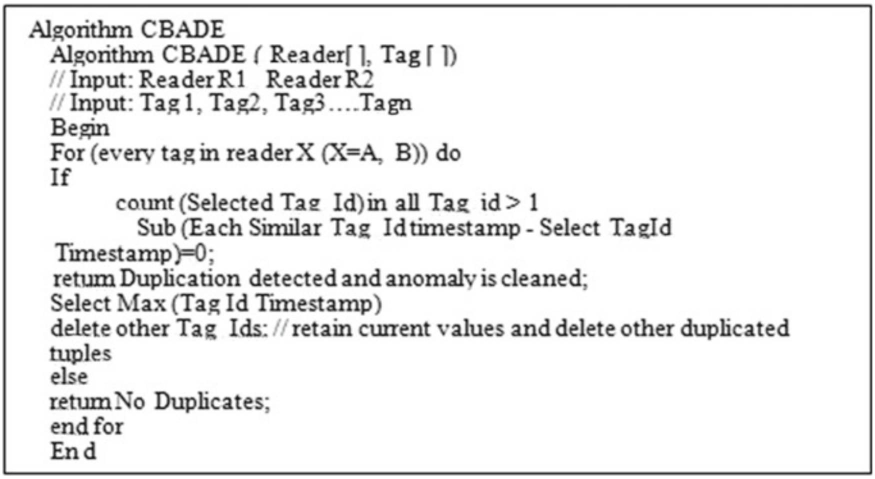
Case study: To test our algorithm the RFID readers are deployed in various departments in the hospital environment. The observed by the readers consist of anomalies. To clean the anomaly duplication the proposed algorithm is CBADE. The following table depicts the sample data observed by the RFID reader. The location assigned to the patient Hari is General ward (102) from 5.00 p.m. to 5.20 p.m. and the time reading is captured every 5 min.

Case	Patient Id	Name	Ward	Reader Id	Date
	Timestamp	Type			
1	P10004	Hari	General Ward	102	10/4/2012
	5:00:00	Normal			
2	P10004	Hari	General Ward	102	10/4/2012
	5:05:00	Normal			
3	P10004	Hari	General Ward	102	10/4/2012
	5:10:00	Normal			
4	P10004	Hari	General Ward	102	10/4/2012
	5:15:00	Normal			

Case	Patient Id	Name	Ward	Reader Id	Date
	Timestamp	Type			
5	P10004	Hari	General Ward	102	10/4/2012
	5:20:00	Normal			
6	P10004	Hari	Infant Ward	105	10/4/2012
	5:20:00	Adjacent			
Parallel					
7	P10004	Hari	General Ward	102	10/4/2012
	5:25:00	Normal			
8	P10004	Hari	General Ward	102	10/4/2012
	5:30:00	Normal			

In the above 1, 2, 3, 4, 5, 7, 8 tuples reader observed readings for P10004 are in the Allotted Location at the Allotted Time and Date. So this tuple is treated as normal and in 6th case the reading shows he is found to be in infant ward (105) which is an adjacent cell of General Ward (102) and P10004 is read by both 105 and allotted location (102) at the same time and date. Therefore the sixth case is treated as adjacent redundant (duplicate data) and with the human intervention based on business rule the duplicated data can be deleted.

7.1 Proposed Algorithm: CBADE



```
Algorithm CBADE
  Algorithm CBADE ( Reader[ ], Tag [ ])
  // Input: Reader R1   Reader R2
  // Input: Tag1, Tag2, Tag3....Tagn
  Begin
  For (every tag in reader X (X=A,  B)) do
  If
          count (Selected Tag  Id)in all Tag  id > 1
             Sub (Each Similar  Tag  Idtimestamp - Select  TagId
   Timestamp)=0;
   return Duplication  detected and anomaly is cleaned;
  Select Max (Tag Id Timestamp)
  delete other Tag  Ids: // retain current values and delete other duplicated
  tuples
  else
  return No  Duplicates;
  end for
  En d
```

7.2 Implementation

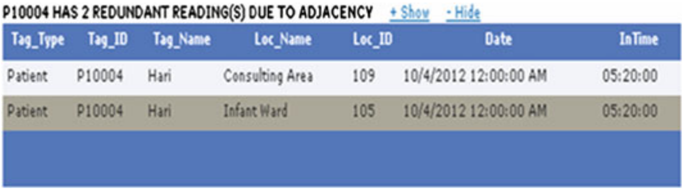
```
SqlDataReader dg15;
SqlCommand cm15;
d1 = "Update RFID_READING set Status ='Adjacent Parallel'
where EXISTS (SELECT * FROM Temp WHERE Temp.Time=
RFID_READING.Time and
Temp.Date= RFID_READING.Date ) and Tag_ID='" +
TagIDDropDown.SelectedValue + "' and Status='Adjacent'";
cm15 = new SqlCommand(d1, con);
dg15 = cm15.ExecuteReader();
dg15.Close();

protected void LinkButton11_Click(object sender, EventArgs e)
   {
Table = new DataTable();
string str = null, tag_ID = null;
tag_ID = TagIDDropDown.SelectedValue;
try
{
        str = "select * from RFID_READING where Tag_ID='" +
TagIDDropDown.SelectedValue + "' and
(Status='Adjacent Parallel' or Status='Crossover Parallel') order by Time";
SqlCommand cmd = new SqlCommand(str, con);
SqlDataAdapter ada = new SqlDataAdapter(str, con);
ada.Fill(Table);
TableGridView5.DataSource = Table;
TableGridView5.DataBind();
}
catch (IndexOutOfRangeException ex)
{
WebMsgBox.Show("Error:" + ex.Message);            }
```

8 Sample Output

Figure 6 depicts the output of the proposed algorithm CBADE. The allotted location of the person Hari is general ward and the schedule time is 5.00 to 5.30 p.m. Figure 7 depict the redundant data in RFID readings.

First Case: 109 is an adjacent cell and P10004 is read by both 109 and alloted location at the same time and date.

Second Case: 105 is an adjacent cell and P10004 is read by both 109 and alloted location at the same time and date.



P10004 HAS 2 REDUNDANT READING(S) DUE TO ADJACENCY + Show - Hide

Tag_Type	Tag_ID	Tag_Name	Loc_Name	Loc_ID	Date	InTime
Patient	P10004	Hari	Consulting Area	109	10/4/2012 12:00:00 AM	05:20:00
Patient	P10004	Hari	Infant Ward	105	10/4/2012 12:00:00 AM	05:20:00

Fig. 6 Output of the CBADE algorithm

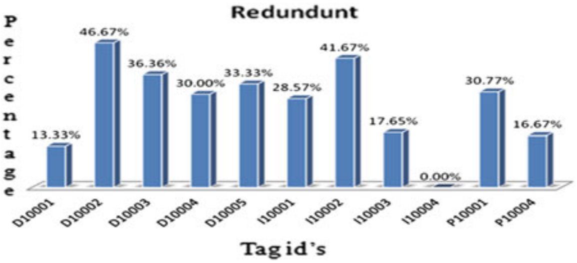


Fig. 7 Chart depict the redundant data in RFID readings

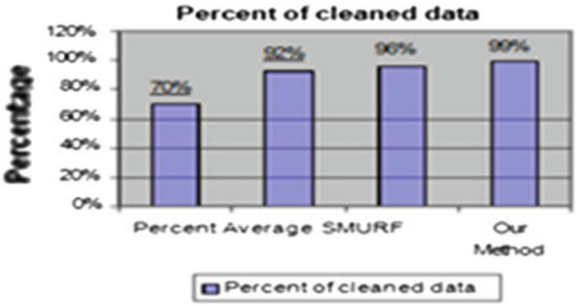


Fig. 8 Chart depict the redundant data in RFID readings

9 Comparative Studies

To show the practical relevance of our method for data duplicates elimination experimental evaluations have been carried out. We analyze the performance of different cleaning schemes. Our evaluation metric is the results of cleaned data. We simulate 1000 samples with 200 wrong Data and see results of cleaning algorithm.10 wrong data is remained. We compare our method with SMURF and other popular strategies depicted in Fig. 8. The average errors are calculated based on following equation

(falsenegatives+
falsepositives)/
NumTags

10 Conclusion

RFID plays an essential role in all the subdomains of the applications in health care applications. Necessary information is provided to patients and hospital staffs by recording and processing the medical data produced in each step, and ultimately this system can be used anytime and anywhere by managing the record of individual health-information. Additionally, these services provide for patients with medical information and guidance of hospital through I/o interfaces by reading individual ID from the medical card of patient with RFID reader and searching them in DB of control center. The effectiveness in cleaning the RFID data in healthcare sectors remains a concern, even though a number of literary works are available. To a maximum, the dirty data that are read because of these errors may even leads to patients' death. The errors need to be cleansed in an effective manner before they are subjected to warehousing. Current solutions to correct missed readings usually use time window filtering. A serious issue is that a single static window size cannot compensate for missed readings while capturing the dynamics of tag motion. An adaptive time window filtering (SMURF) cannot deal with the condition that tags are always moving. In this paper, we have proposed an improved algorithm CBADE to clean the anomaly duplication and the experimental result has proved our algorithm predicts and removes the data duplication in an effective manner compared to the existing works. Thus it will pave the way for an effective means of data warehousing system that will keep the RFID data safe for future mining.

References

1. Shepard S (2005) RFID Radio frequency identification. McGraw-Hall, New York
2. Durresi A, Merkoci A, Durresi M, Barolli L (2007) Integrated biomedical system for ubiquitous health monitoring. NbiS 4658(1):397–405
3. U.S. Government Accountability Office (2005) Radio frequency identification technology in the Federal Government, 441 G Street NW. Room LM Washington, DC 20548
4. Zhang C, Chen Y (2011) Application oriented data cleaning for RFID middleware. IEEE Trans Commun 59(1):159–168
5. Floerkemeier C, Lampe M (2004) Issues with RFID usage in ubiquitous computing applications. Pervasive 3001:188–193
6. McGlynn EA, Asch SM, Adams J et al (2003) The quality of health care delivered to adults in the United States. N Engl J Med 348(26):2635–2645
7. Shoewu O, Badejo O (2006) Radio frequency identification technology: development, application, and security issues. Pacific J Sci Technol 7(2):144–152

8. Aragonés J, Martínez-Ballesté A, Solanas A (2007) A brief survey on RFID privacy and security. In: Proceedings of the word congress on engineering WCE07. IAENG, pp 1488–1493
9. Uddin MJ, Ibrahimy MI, Reaz MBI, Nordin AN (2009) Design and application of radio frequency identification systems. Eur J Res 33(3):438–453. ISSN 1450–216X
10. Anny Leema A, Hemalatha M An effective and adaptive data cleaning technique for colossal RFID data sets in healthcare. WSEAS Trans Inf Sci Appl
11. Chawathe SS, Krishnamurthy V, Ramachandran S, Sarma S (2004) Managing RFID data. In: Proceedings of the 30th VLDB conference, pp 1189–1195
12. Rao J, Doraiswamy D, Thakkar H, Colby LS (2006) A deferred cleansing method for RFID data analytics. In: Proceedings of the 32nd VLDB conference, pp 175–186
13. Wicks AM, Visich JK, Li S (2006) Radio frequency identification applications in hospital environments. Hosp Top 84(3):3–9

Part IX
Workshops: Fourth Workshop on Applications of Graph Theory in Wireless Ad hoc Networks and Sensor Networks (GRAPH-HOC - 2012)

Energy Aware Multipath Routing Protocol for Wireless Sensor Networks

Suraj Sharma, Pratik Agarwal and Sanjay Kumar Jena

Abstract Wireless Sensor Networks (WSNs) are made of sensor nodes with restricted battery life and transmission capability. In this paper we propose an energy efficient multipath routing algorithm in WSN. This protocol is designed to improve the lifetime, latency and reliability through discovering multiple paths from the source to the destination. It has a sink initiated Route Discovery process with the location information of the source known to the sink. There are two types of nodes which are used here one is primary and the other is alternate. At the end of the route formation one primary path and a number of alternate paths are built and all nodes except the primary are put to sleep mode which helps us to save energy and generate a collision free path, the primary path is used to transmit the data from source to the sink and if the route disrupts, the next best alternate route is used for the purpose and if no path exists between the source and destination then process starts from the beginning. Further, we analyze how the proposed protocol overcomes the drawback of the existing protocols.

Keywords Wireless sensor networks · Multipath routing · Energy-efficiency

S. Sharma · P. Agarwal · S. K. Jena
Department of Computer Science, National Institute of Technology, Rourkela, India
e-mail: suraj.atnitrkl@gmail.com

P. Agarwal (✉)
e-mail: pratikagarwal.nitr@gmail.com

S. K. Jena
e-mail: skjena@nitrkl.ac.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_73,

1 Introduction

Wireless Sensor Network is a collection of largely deployed sensor nodes with limited computational, transmission capability and power supply, the density of nodes is high so the communication takes place through large number of hops. It is also used in health care, area monitoring, air pollution monitoring, greenhouse monitoring, landslide detection and many other applications.

Based on literature routing protocols are categorized into Structure based and Operation based [1]. The structure based can be classified into flat, hierarchical and location based. The operation based routing protocols are; negotiation based, query based, QoS based, multipath based, and coherent based protocols.

As we know that Wireless Sensor Network are built from sensor nodes with very less transmission and computation capability and lifetime, so to use single path routing protocol is very inefficient regarding energy, which decreases the reliability of the network, which also affects the QoS, in this regard, to increase the reliability and QoS and make the network energy efficient it is fruitful to use multipath routing [2] protocol, as the name suggest there are multiple path between the source and the sink, if one route fails there is alternative paths to reach the destination. Following are the advantages of multipath routing protocol:

Reliability: It increases the reliability of the network making it fault tolerant as there are multiple paths to reach the destination.

Load Balancing: Multiple paths are built to send the data which helps in reduce the network traffic congestion and distribute the load among the sensor nodes to relay data.

Quality of Service: We can improve the network throughput, end to end latency, and data delivery rate as the time critical data can be sent through shorter path decreasing the delay and non time critical data can be sent through longer paths.

There are certain energy efficient multipath routing algorithms such as Energy Efficient Multipath Routing [3], Energy Efficient Collision Aware Multipath Routing [4], Maximally Radio Disjoint Multipath Routing [5], and Low Interference Energy Efficient Multipath Routing Protocol [6]. These all energy efficient protocols either are based on collision aware with least interference or load balancing.

In this paper we consider these two aspects i.e., collision avoidance and load balancing, of existing Multipath Routing Protocols, in addition we proposed a new Energy Efficient Multipath Routing Protocol which uses non-flooding based method to generate multiple paths between source and sink.

The rest of the paper is organized as follows: Related work is discussed in Sect. 2. Section 3 describes Motivation and System assumption of the proposed protocol. Algorithm and proposed protocol are briefly discussed in Sect. 4. In Sect. 5 we did detail analysis of the proposed protocol with respect to existing protocols followed by Conclusion and References.

2 Related Work

As we discussed energy aware protocols have been used two method; collision aware and load balancing. Collision aware means that when one node is sending data through a route then other nodes who are not in the route must be aware that a data transmission is going on and should not interfere in between.

Load balancing simply means to balance the network load over the whole network so as the pressure doesn't come on a single path alone, that creating multiple paths and sending the parts of data through different paths so that the load is balanced among the nodes.

LIEMRO (Low Interference Energy Efficient Multipath Routing) [6] was basically designed to improve the packet delivery ratio, latency, and lifetime through multiple interference-minimized node disjoint paths in addition it includes a load balancing algorithm to distribute the source node traffic over multiple paths based on relative quality of each path. Reference [3] is a distributed, scalable, localized multipath search algorithm, to discover multiple node disjoint paths between sink and the source nodes, it also used load balancing algorithm that distributes the traffic over the multiple paths discovered. It extracts the path diversity by distributing the network traffic over different node disjoint paths. Reference [4] is an energy efficient node disjoint multiple path routing algorithm, with the aid of node position information, it finds out two collision free routes with constrained and power adjusted flooding and then transmit the data with minimum power required through power control component of the protocol. In [5] the main objective was to provide necessary bandwidth to multimedia applications through non interfering paths. It is an incremental approach, only one path is built at a time and additional paths are built when required typically in case of network congestion or bandwidth shortage. Energy saving is done by putting the interfering nodes to passive state that is sleep state and after going to passive state they will not take part in any routing process. AOMDV-Inspired Multipath Routing Protocol [7] is based on the multipath version of AODV which uses cross layer information to achieve energy efficient and low latency communication in wireless sensor networks. AOMDV tries to discover all node disjoint paths in between the source and the destination but here the hop count, and the first hop is taken into consideration, the sink allows an additional path only if the first hop is different and the hop count is same else if a path with lower hop count is found all the previous paths are deleted. It does not provide any load distribution mechanism to distribute the load over the network.

3 Proposed Protocol

We proposed a protocol based on the multipath scheme where multiple route exist between each source and the sink. In this section we discuss system model, assumptions and working principle for the proposed protocol.

3.1 Network Model and System Assumption

There are *n* number of sensor nodes and a sink node in the network. After deployment the nodes will be stationary. We consider the sink node is in the middle of the network and static in nature. Sink possesses unlimited computation, memory, and battery power. Each node knows their position and sink node contains the id and position of each node in the network. Sensor nodes are densely deployed and all are homogeneous. Each nodes communication range are identical and predefined.

3.2 Energy Aware Multipath Routing

In the paper we proposed a routing algorithm which creates partially node disjoint paths.The protocols avoid flooding and takes the benefit of both load balancing and collision aware mechanism for energy conservation. Proactive routing protocol is preferred for the static network, but it is not advisable for the resource constrained sensor network. So we construct the route between source and sink when actually sink need the data from a particular source node. With this requirements we design two multipath routing algorithms for WSN. They mainly consists of three phases; Neighbor Discovery, Multipath Construction, and Data Transmission.

1. Neighbor Discovery

In this phase every node broadcasts a control packet contains their node ID, residual energy, and the location and wait for the neighbor discovery control packets from the nodes of its range to find the neighbor nodes. After the neighbor discovery phase each node finds its neighbor nodes. Now every node has the partial view of the network. Algorithm 1:

2. Multipath Construction

After the Neighbor Discovery phase each node possesses their neighbor information. Then Multipath Construction phase starts, we assume that the source location is known to the sink and based on the location of the source the sink starts the route request process,in this the main concept is that, there are two type of nodes (Primary and Alternate), a node is Primary if it is in the primary path from source to sink else if it is the part of any alternate path then it is the Alternate node. As described in the Algorithm 1, if node is Primary then it will find two paths to the source primary path and alternate path, the primary path is built with the best possible neighbor [having the maximum signal strength and the minimum Location Factor (LF)] and the alternate path will be constructed with the next best neighbor (having the next maximum signal strength and the next minimum Location Factor (LF) after the primary path node) and if a node is Alternate node then it will find the node with maximum signal strength and will prefer a Primary node if possible, this is done just to converge the path else the path can diverge from its direction toward the source,

Algorithm 1 **Multipath Construction**

```
repeat
    if ($node == sinknode$) then
        $FindPrimaryPath()$;
        $FindAlternatePath()$;
    else if ($node == Primary$) then
        $FindPrimaryPath()$;
        $FindAlternatePath()$;
    else if ($node == Alternate$) then
        $FindPrimaryPath()$;
    end if
until ($node \neq Source$)

procedure $FindPrimaryPath()$
    if ($node == Primary$) then
        $Broadcast$ PRIMARY;
        Search for the best node;
        $node \leftarrow Primary$;
    end if
    if ($node == Alternate$) then
        $Broadcast$ ALTERNATE;
        Search for the best node and prefer Primary;
        if ($node \neq Primary$) then
            $node \leftarrow Alternate$;
        end if
    end if
end procedure

procedure $FindAlternatePath()$
    if $node == primary$ then
        Search for the next best path node accept Primary;
        if (($node \neq Primary$)&&($node \neq Alternate$)) then
            $node \leftarrow Alternate$;
        end if
    end if
    if ($node == Alternate$) then
        $Exit()$;
    end if
end procedure
```

here Location Factor $LF = L_a - L_b$ where L_a is the location of the source and L_b is the location of neighbor node.

Here it is an incremental approach from the sink to the source, first the sink node which is itself a Primary node finds two paths first primary and the other is alternate by this there are now two more nodes Primary and Alternate in the list, and using these two nodes it can go forward with multipath construction. The second checking is done for if the node is primary then again it will find two more nodes Primary and Alternate and add the nodes to the list towards the source and the third checking is for whether the node is Alternate. It will find a Primary node which is having the best signal strength and the minimum location factor, and this full process will go on till reaches the source. Figure 1 illustrated about the multipath construction, the strong arrow line shows the primary and dotted arrow line shows the alternate path. Meanwhile each node associated with path makes a reverse link toward the sink so that the data can transmit to the sink.

FindPrimaryPath(): This function is called by both Primary and Alternate nodes, so if the node is Primary it will broadcast a PRIMARY message with its id stating that

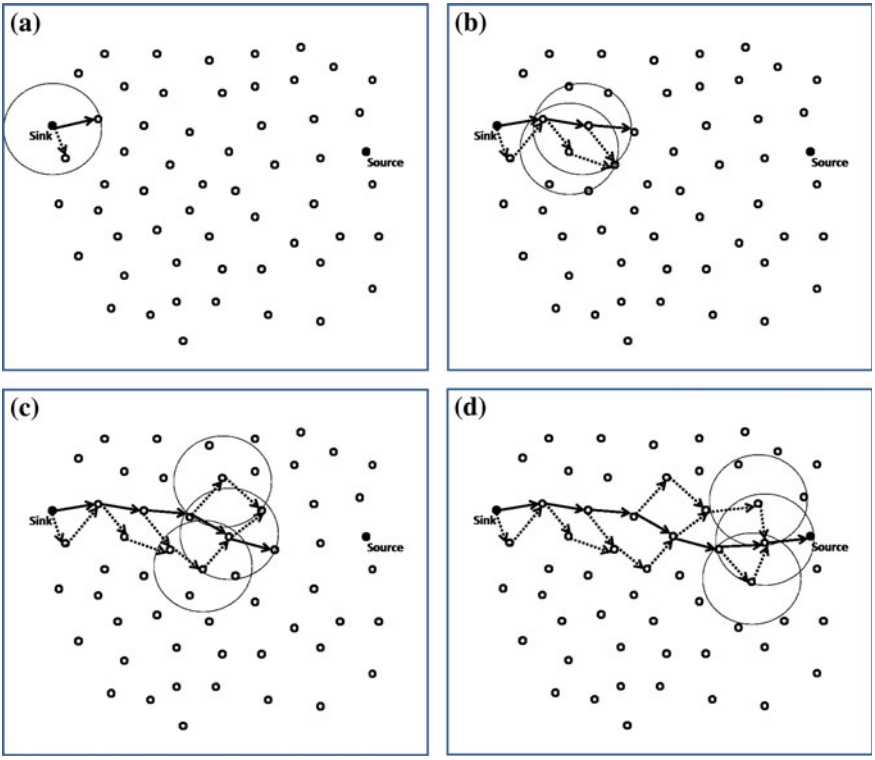


Fig. 1 Multipath construction steps in (**a**–**d**)

it is the Primary node and search the best possible node, and if the node is Alternate it will broadcast a ALTERNATE message with its id stating that it is the Alternate node and find the best possible node towards the source and will prefer the Primary node if possible, so that the path converge instead of diverge.
FindAlternatePath(): This function is called only by the Primary nodes for finding an alternate path towards the source. It find the next best node which is called Alternate node and add it in its path.

In the algorithm all nodes except the Primary nodes are put to sleep mode, and as at a time we have only one primary path so we can avoid interference from other paths to reduce the collision. Both of these factors help to save energy. If the primary path disrupts the protocol selects the alternate path to transmit data, and if all path disrupts and no path is left between the source and sink then again starts from the Neighbor Discovery phase.

3. Data Transmission

After the route discovery by the multipath construction phase data transmission takes place between source and sink. The primary and alternate paths are available, but the data transmitted only over the primary path. Source utilize alternate path when primary path is not available. Rest of the nodes that are not in the active path will go to sleep mode to conserve energy. If there will be no path exist between source and sink, the process of route discovery starts.

4 Theoretical Analysis

In comparison to EECA [4] in the proposed algorithm we have overcome the problem of flooding , secondly here two active paths are built where to avoid interference there is an constraint that nodes of two different paths should be at least R distance apart but we are constructing one primary path and multiple alternate paths so there is no problem of interference, thirdly the proposed approach is a sink initiated route discovery process and there is no overhead of route reply to the route request.

In comparison to LIEMRO [6] our proposed algorithm has a sink initiated route discovery process so there is no overhead of route reply, secondly as we are considering only single primary path and multiple alternate path so we don't have the overhead of considering the interference of all nodes, thirdly at a time we are establishing one primary and multiple alternative path, but in the above algorithm at a time only one path is constructed and for the second path again the process starts from the beginning.

In comparison to Energy Efficient Multipath Routing for Wireless Sensor Networks [3] our proposed algorithms uses broadcast method that too once only in the neighbor discovery phase then it is a multicast process of route construction secondly in the above algorithm there are multiple paths constructed in which interference between nodes are not taken into consideration but in our algorithm only the nodes of the primary path are active and rest are in sleep mode so there is no interference and with this we can increase the network lifetime and make the system more energy efficient.

In comparison to Maximally Radio Disjoint Multipath Routing for Wireless Sensor Network [5] the main drawback that we have overcome on this protocol is that in the above algorithm they used flooding in the route discovery process from the sink to the source for building the route but we used multicasting in the route construction phase.

In comparison to AOMDV-Inspired Multipath Routing Protocol [7] we use a non-flooding route discovery technique, secondly we have a sink initiated route discovery with no overhead of route reply, third we are not keeping any overhead of the neighbor node time information.

5 Conclusion

In this paper we have proposed an energy efficient multipath routing protocol for Wireless Sensor Networks. The detailed analysis shows that how the proposed protocol overcomes the drawback of the existing protocols. We summarized the

Table 1 Comparison of proposed and existing protocols

Algorithm	Route request message	Path disjointedness	Source initiated	Sink initiated	Collision avoidance	Load balancing
EECA [4]	Flooding	Node disjoint	√	×	√	×
LIEMRO [6]	Multicasting	Node disjoint	√	×	×	√
MR2 [5]	Flooding	Node disjoint	√	×	√	√
EEMR [3]	Flooding	Node disjoint	√	×	×	√
AOMDV-IMRP [7]	Flooding	Node disjoint	√	×	×	×
Proposed protocol	Multicasting	Partially node disjoint	×	√	√	√

comparison of the existing and proposed protocol in Table 1. The proposed protocol generates the routes (Primary and Alternate) between source and destination without using flooding and also provides collision aware and load balancing methods which makes it energy-efficient and helps to increase the network lifetime.

References

1. Al-karaki JN, Kamal AE (2004) Routing techniques in wireless sensor networks: a survey. IEEE Wirel Commun 11(6):6–28
2. Radi M, Dezfouli B, Abu Bakar K, Lee M (2012) Multipath routing in wireless sensor networks: survey and research challenges. MDPI Sens 12(1):650–685
3. Lu YM (2007) Wong VWS (2007) An energy-efficient multipath routing protocol for wireless sensor networks. Int J Commun Syst 20(7):747–766
4. Wang Z, Bulut E, Szymanski BK (2009) Energy efficient collision aware multipath routing for wireless sensor networks. In: Proceedings of the (2009) IEEE international conference on communications. IEEE Press, New York, pp 91–95
5. Maimour M (2008) Maximally radio-disjoint multipath routing for wireless multimedia sensor networks. In: Proceedings of the 4th ACM workshop on wireless multimedia networking and performance modelling. ACM, New York, pp 26–31
6. Radi M, Dezfouli B, Abd Razak S, Abu Bakar K (2010) Liemro: a low-interference energy-efficient multipath routing protocol for improving qos in event-based wireless sensor networks. In: Proceedings of the (2010) fourth international conference on sensor technologies and applications. IEEE Computer Society, Washington DC, pp 551–557
7. Hurni P, Braun T (2008) Energy-efficient multi-path routing in wireless sensor networks. In: Proceedings of the 7th international conference on ad-hoc, mobile and wireless networks. Springer, Berlin, pp 72–85

Virtual Classroom for E: Education in Rural Areas

Vimal Upadhyay, Mukesh Chand and Piyush Chaudhary

Abstract E—Education is an emerging practice, includes various processes and assimilates government responsibilities at different levels and ventures of public and rural sectors. These kinds of facilitation of different area of rural India extremely helpful to establish good education in developing countries having inadequate resources with surplus manpower and deficit finance. Wireless Cloud computing, a collection of multiple wireless computer interfaces through a digital network. It is an excellent model for convenient which is based on demand network access to share the configurable computing resources. Wireless Cloud is used as a resources configuration for rural education in India as presented in this paper. The interoperability between different components of the Wireless Cloud computing represents each segment of development of rural education such as e-learning, virtual classrooms, e-lectures, e-study centre through proper balancing of teacher-student relational input of rural communication. Simultaneously e-schooling architecture especially for rural education is also presented through Wireless Cloud computing taking each block, district and state as distribution of a standard education equally. An interoperability parameters has also been presented by using three important and basic elements of computing i.e. memory, bandwidth and processing capability essential for the communicative link between the teacher as sender and the student as receptor specially in India. A QualNet based parallel computing simulation approach has also been presented for implementation.

V. Upadhyay
Research Scholar IIIT, Allahabad, UP, India
e-mail: vimalupadhyay2002@gmail.com

M. Chand (✉)
M-Tech Scholar, SITM, Rewari, Haryana, India
e-mail: mukeshchand@gmail.com

P. Chaudhary
SMEC, Neemrana, Alwar, Raj, India
e-mail: piyushjbad@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_74,

Keywords Wireless Cloud computing · E—learning · Rural education · Virtual classroom

1 Introduction

Wireless Cloud computing means the use of multiple wireless computers through a digital network though they work as an individual single computer. Wireless Cloud computing is a model for enabling heterogeneous resources that can be rapidly conjugated and released with remote management effort or service provider interaction [1]. From the view of rural education factors the Wireless Cloud computer is used as the national integrates system. The interoperability between deferent components of the Wireless Cloud computing as the clusters or grid computing are worked as the individual interfaces at the rural education centres in villages and state levels universities which are interconnected with the each other [3]. The concepts of the collection of grid conserve with the system as the integrated way. The coefficient of interoperability of the integration factors, the grid are represented as the district individualities and the cluster computer are interfaced as the interconnectivity between the different virtual classrooms and the e-educational centres in villages [4]. The interoperability factors in the Wireless Cloud, grid or clusters are merged through the neural network and calculus integration in every level of the rural e–learning especially in India where the literacy rate in rural areas are not in count [2]. In the higher architecture the Wireless Cloud computing is the central level of horizontal integration among the each deferent virtual college or virtual university centres in zonal and country areas. The flow of the data at the state or rural level is done by the cluster computers as the differentiable components from e-learning like rural educations, virtual classrooms, virtual schooling etc. [2] so, the each rural education centres or virtual schools are connected through the Wireless Cloud, cluster in upper level and grid computing with each other in every districts and zone by the interoperability factors between the each components of the system in e-learning at the country to central or international integrated level simultaneously [5]. The Wireless Cloud computing as the whole integrated system with the interfaces at the village level to international level by the Wireless Cloud computing and its components for the rural educations in deferent domains like e-learning of health, agriculture etc with the help of the interoperability factors at the e-college and e-university level centres in horizontal way [6]. District as the grid and state as the cluster with country as the Wireless Cloud for the international aspects are defined in the integrated computing with proper interconnectivity in the field of research and educations at the rural and country areas [3].

2 Virtual Classroom and University

In the environment the European universities produce useful workforce for standardizing which it is also be needed to introduce in India especially in rural areas. The teaching materials and lecture notes are standardized at world class of labor. On one hand Universities, which cannot develop any solution for these standards, would not be able to setup their institutional settings up to the latest trends as required. On the other hand the using of latest technology in education helps to make the knowledge transfer easier. For that reason it is essential to consider for higher educational and research institutions, what kind of technology should be applicable in education, and how it will benefit all the technical improvement without losing the attention of students.

To account not only for structure and the state-of-affairs but also for development of the activity system(s), focus is placed on tensions that reside within and between activities and which trigger processes of development. Angstrom denote these tensions as contradictions. If changes are made in an activity system, e.g. to one node, tensions are likely to occur and changes are likely to happen in other nodes as the system develops to cope with the contradictions. The theory of activity systems is thus helpful in explaining why activity systems develop as they do, to predict how activity systems are likely to develop in response to certain change, and—to some extent—how change in a certain direction might be prevented or encouraged. In the case of the community of master course learners, any change which results in better learning outcomes on part of the students, are desirable. Changes which do not worsen learning outcomes and which simultaneously offer other advantages (such as cost savings in the interrelated activity system of the school), are also desirable. The relation between virtual classroom, curriculum and students which come under the e-learning system is represented in the figure below (Fig. 1):

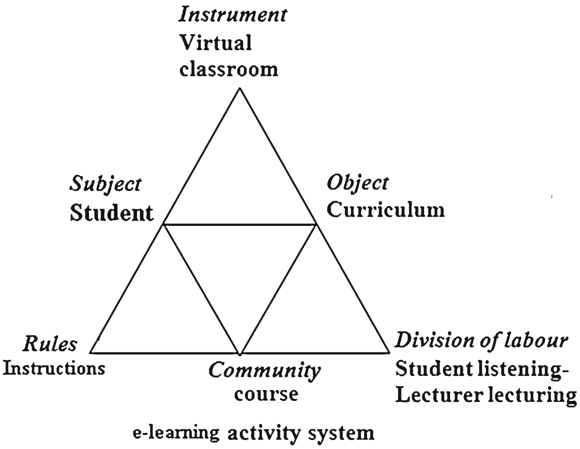


Fig. 1 Pyramid structure of e-education system

Virtual Classrooms could be an effective and efficient solution for people at remote distance where local educational facility is very poor in quality. For the universities it will be efficient as they can transfer their knowledge towards the remote costumers through virtual classroom technique. The global cooperation among the universities also indicates the necessity of scalability, efficiency, portability of knowledge, transfer methodologies and frameworks such as Virtual College, virtual university. Virtual Learning Environments in education claim a massive, safe IT architecture, excellent professional skills for setting up the infrastructure. This commitment helps to bring this new virtual dimension into our "everyday life" with the establishment of a new operational setup. These Virtual Universities will be capable of transforming the educational activities of "normal" universities into this new technology and make the higher educational and research institutions ready to face the problems as like other. A Virtual University—as a part of the complete institution—will be responsible for distribution and implementation of the new technologies to the remote citizen.

3 e-Lecture Broadcasting Through Wireless Cloud Computing

Educators can adapt the Wireless Cloud computing to distribute their courses, tutorials and material. For example, after applying the most appropriate Wireless Cloud network, they can send educational materials to the remote areas. The new materials can then be provided to students on internet and made accessible. Initially the students only require the computer which will be provided by the authority and to connect to the educational Wireless Cloud network. Once again, a virtual classroom will be prepared by one educator who can be shared, reused as well as improved by other educators (Fig. 2).

The virtual classroom is fully self-contained: the Wireless Cloud connection is needed to run the video lecture series. The interaction between the student and the teacher is fulfilled by the Wireless Cloud connectivity. The educator can retrieve the Wireless Cloud connection used by the students (or connect to the teacher lecture hall) and checks the video is properly being transmitted or not. The collaboration capabilities of the Wireless Cloud communication open new perspectives in distributed e-learning. The educator can connect the lecture hall from any location where Wireless Cloud connection is stable. Teacher can see and update their educational setup environments and guide them remotely. Problem solving with collaboration will be also possible and can be used as a support of e-learning. So this governing force should have enough authority to introduce Virtual Classrooms in the e-education which is not very easy task. Setting up all the hardware and software architecture, convincing the different departments about their respective benefits from this new system, founding an uninterrupted network service to maintain its functionality, capabilities as well as removing the administrative burdens for the proposed Virtual Learning Environment, are very challenging tasks, which needs a strong administrative support not only from the leaders of universities but also from the governing authority as well.

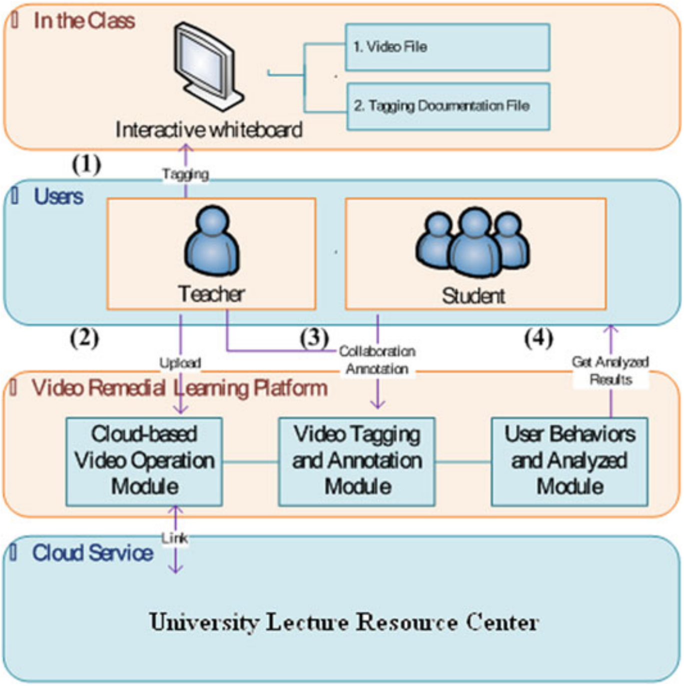

In the Class
Interactive whiteboard
1. Video File
2. Tagging Documentation File
(1)
Tagging
Users
Teacher
Student
(2)
Upload
(3)
Collaboration Annotation
(4)
Get Analyzed Results
Video Remedial Learning Platform
Cloud-based Video Operation Module
Video Tagging and Annotation Module
User Behaviors and Analyzed Module
Link
Cloud Service
University Lecture Resource Center

Fig. 2 Uploading and video streaming of e-lecture series through Wireless Cloud orientation

4 e-Educational Centre Connectivity

The concept of virtual universities is spreading very fast among the developing countries where the ICT industry is growing up exponentially especially in India. According to the ideas of virtual university or centre is not real university but whose primary goal is to teach the students and crown them a degree after completion of all requirements. A Virtual classroom has to support the networking and teaching facilities as their real university, which includes the following tasks. In India individual requirement based e-learning systems are now widely spreaded and most of the university or colleges offer e-learning courses. As in the distance learning scenario the group-based systems are hardly implemented. These e-learning systems are known as “virtual classroom training systems”.

The basic element in this protocol is the Wireless Cloud Node, also referred to as the Wireless Cloud host, which roams among its university network and other remote networks. Mobility is achieved in part by having a host on the Wireless Cloud node’s university network, called the university Agent, which is responsible for trapping packets destined to the Wireless Cloud node, and then forwarding them to the present location of the Wireless Cloud node, if it is away from university location. The present location of the Wireless Cloud node is in the form of address.

This care-of address is obtained either directly by the Wireless Cloud node using an external assignment mechanism (i.e., Dynamic Host Configuration Protocol [DHCP]) or from a special node, called a remote Agent, which is present on the remote network and provides mobility services to visiting Wireless Cloud nodes. Packets are forwarded from the university Agent to the care-of address by encapsulating and tunnelling them 5 address at the end of the tunnel is that of the remote Agent, the packets are decapsulated by the Remote Agent and forwarded by link-level mechanisms locally to the Wireless Cloud node (Fig. 3).

5 Virtual Classrooms Through Wireless Connectivity

E Education services in India have been enormously admired since last few years because it has very large target segments of virtual schooling, college and university bodies. E Education operates at the intersection among Information and Communication Technology and educational processes. E Education could be alienated into three overlapping spheres: E-learning, E-lectures and E-examinations. The end objective of this paper is to manifest and assert wireless connection strength in E Education application manly in virtual classroom in rural areas of India. Wireless communication provides efficient techniques for educational bodies to provide the video lectures quickly at country areas. The wireless video transmission generates the facility at the rural area. The e-learning enables common standard educational systems in those areas where the primary education is still beyond reach. The main function of VTA (virtual teaching application) module is the collaborative video annotation. And the main motto is to increase the interaction between users, and to promote collaborative

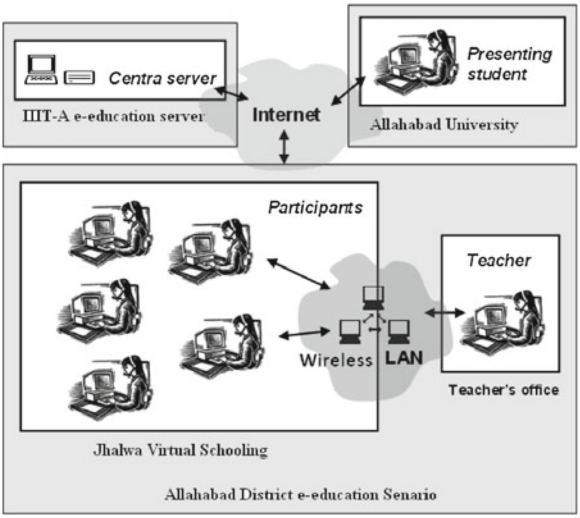


Fig. 3 Allahabad District e-education system

learning through sharing the video notes. The teacher directly inserting course conception during in-class recording process, then upload the video through the platform after the video output, and you can make comments at the platform to share study notes. According to the browsing path through user's behaviour to provide an efficient searching and learning for users by convert the tagging content into meaningful knowledge structure. The purposes of E learning system are to establishing a good and standard education system and have seamless coordination between university authorities and virtual classroom coordinators in rural areas especially in India. The utilization of ICT may combine all three different sectors to support the development and management of e-education system (Fig. 4).

E-learning is a way of management of any education system supervised by an administrator. In any educational system the administration may be governed by third party organization or a centralized authority. The term E is basically the responsibility of a electronic which includes each and every processes performed by electronically. The main activity of the e-education is to control the education of different aspects for example e-schooling, e-college, e-lectures etc. All these activities of e-learning are now maintained by using ICT efficiently. The transformation of the flow of educational from the conventional setup to modern structure of ICT is helpful for E education. The uses of ICT in governmental activities have provided a new idea of learning system known as E learning (Fig. 5).

6 Virtual Schooling by Wireless Cloud Networking

The centralized control server based deployment of the Wireless Cloud Computing Environments enables new scenarios in the field of e-learning. With Wireless Cloud computing, it becomes possible for educators to communicate the remote area where

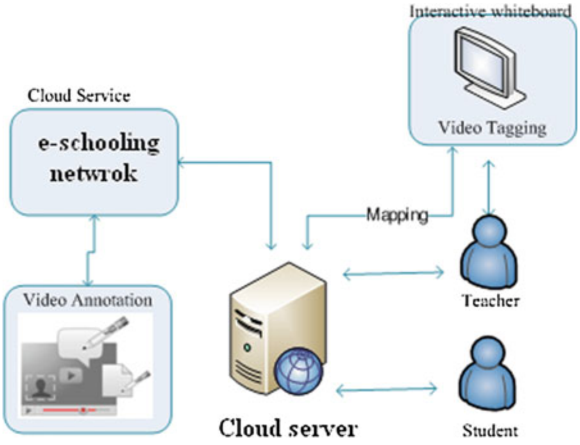


Fig. 4 E-lecture and video annotation module for wireless cloud server

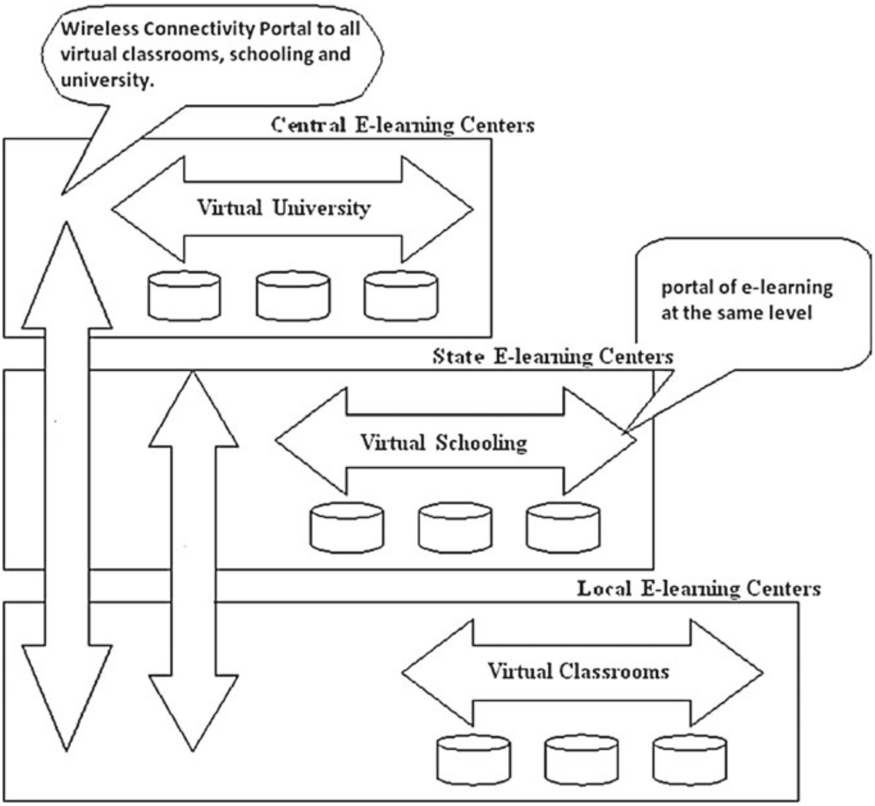

Wireless Connectivity Portal to all virtual classrooms, schooling and university.
Central E-learning Centers
Virtual University
portal of e-learning at the same level
State E-learning Centers
Virtual Schooling
Local E-learning Centers
Virtual Classrooms

Fig. 5 Interconnectivity diagram of different level of e-learning

availability of education resources is very poor. These are very easy to create and to distribute among the students. The User Interfaces reduce the complexity of learning environment and keep students away from its steep learning curves. Though an e-learning environment is created by one educator, the Interfaces can be shared, reused as well as improved by other one. Dedicated study material repositories can be centralized the efforts and contributions from the community of educators and help them sharing the gained knowledge by using this new environment. One could can envisage these methods which can be used from primary schools to graduate-level studies. The e-learning itself is available as a public virtual classroom on Wireless Cloud communication. That virtual classroom technique can be used to decentralize the private-virtual collaboration for e-learning environment. Because of Wireless Cloud resources are not free; students might not be able to use their own Wireless Cloud communicators. Virtual classroom provides a mechanism of e-learning that can be delivered by the University information resource manager to researchers, educators and students (Fig. 6).

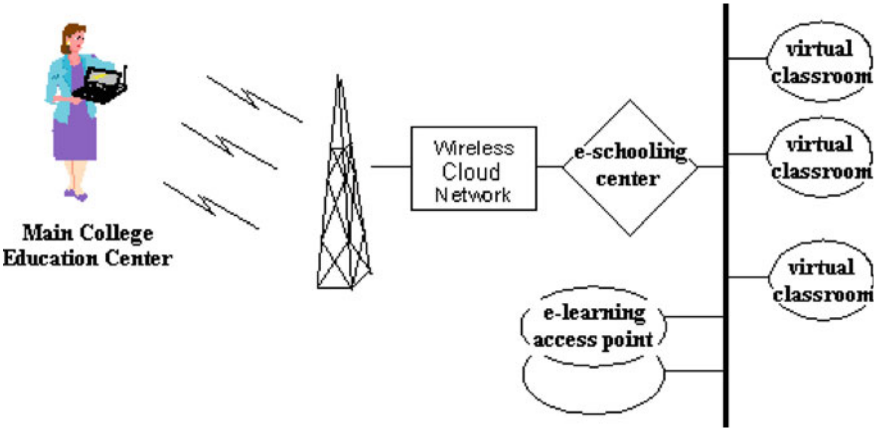


Fig. 6 Enabling wireless virtual classrooms

The overall regulation of e-learning centres may be carried out by using appropriate technology. The third party is Service provider for wireless connectivity to the remote area, which includes all available operations of the E education. It provides an interface between user and deliver system. The main block is wireless connectivity block, which contains various categories of users working in this environment. The user categories may be a student, researcher, teaches. The proposed E learning model covers all important aspect of E lectures in a single model with virtual classrooms. There are Basic wireless Blocks of proposed E education Model. The lowest block is the controlling Block, which regulates the overall functionality of any district or block through efficient learning system node.

7 Virtual Classrooms Through Wireless Connectivity

The need for the Wireless Cloud worker to access mission critical information requires access to corporate databases and Internet/Intranet applications. In addition, convenient and reliable file transfer, integrated messaging, and personalized information delivery allow the Wireless Cloud employee to work at peak productivity levels. Virtual classroom Networks have emerged to provide networking solutions to a growing Wireless Cloud workforce. A Virtual classroom Network allows businesses to provide their Wireless Cloud employees could access the corporate information and applications by connecting them to a public networks, such as the Internet. By using public networks as the communications backbone, a Virtual classroom Network provides a low cost extension to the enterprise, while offering secure access to an open networking environment (Fig. 7).

The tokens allow them to start virtual classrooms for certain duration and use them for their courses and research. Real-life problem for example E education projects

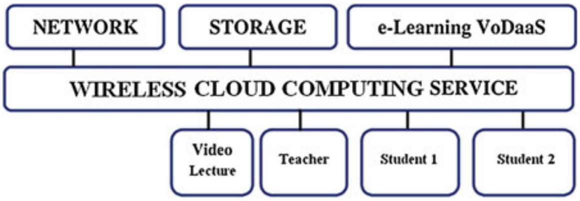


Fig. 7 Virtual e-Research/e-Learning environment

especially in virtual classroom domain consist of huge video streaming, which still needs considerable efforts in terms of bandwidth, processing capacity and memory. It is extremely essential to develop a model with less bandwidth and with minimum cost of transmission on nature of video streaming especially in rural areas. The video transmission through wireless network has been studied in this research work. This study exposed the perspective way to distribute a common and standard educational response all over the rural India. Here it is aimed to show architecture between the virtual study centres at urban a rural areas with the main facilitation at the university or college.

8 End User or Student Level Based Wireless Cloud Networks

According to the browsing path through user's behaviour to further build up personal knowledge ontological conception with ontology. And provide learning path of recommendation, in order to achieve individualized remedial learning. Turn tagging into a kind of linking, and use Wireless Cloud computing algorithm to find out the relation, extracting some beneficial information such as the key points of video and interest of learn (Fig. 8).

A Virtual classroom Network is created when a Wireless Cloud user connects a data terminal to a remote network, either via dial-in or public networks, and establishes a presence equivalent to a direct connection to the e-learning network. Both of these

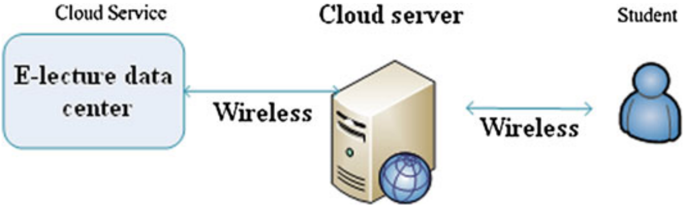


Fig. 8 Student level application of virtual classroom interface

scenarios are undesirable. Altering the Wireless Cloud address would cause the breakdown of existing transport level connections, while the propagation of host routes causes severe scaling problems, especially with an ever-increasing number of Wireless Cloud nodes seeking the host node. Wireless Cloud communication is an Internet industry standard that enhances the Wireless Cloud connectivity to remedy these existing problems and allows transparent routing of Wireless Cloud computing datagram's to Wireless Cloud nodes on the Internet (Fig. 9).

Using the Wireless Cloud computing solution, the Wireless Cloud nodes are always identified by their permanent centre address and their current position in the Internet. In addition to this permanent centre address, the Wireless Cloud node, while away from its university network, is also associated with a temporary care-of address, which provides the information about the current location of attachment in the internet. Wireless Cloud connection makes the goal of location-transparent communications possible by defining a set of mechanisms for Wireless Cloud nodes to acquire a care-of address. It also ensures a means by which packets destined for the Wireless Cloud node (and hence delivered by traditional Wireless Cloud connection mechanisms to the Wireless Cloud node's centre network) are ultimately forwarded to the present location of the Wireless Cloud node, as indicated by its current care-of address. In this study, the video remedial teaching resources will be put into the Wireless Cloud services platform because of the advantages of Wireless Cloud computing service environment, such as reliability, scalability and convenience. In choice of Wireless Cloud computing service environment, but the kind of open and

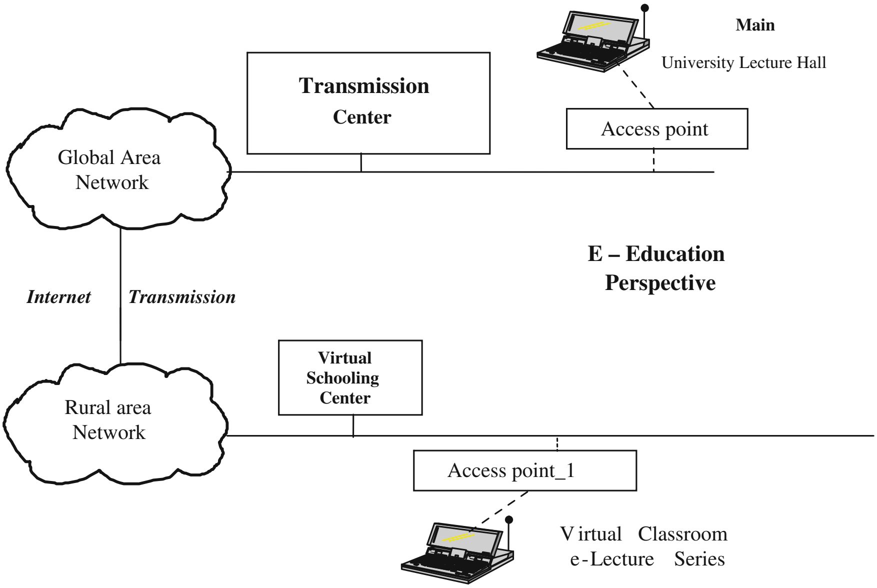


Fig. 9 e—lecture transmission

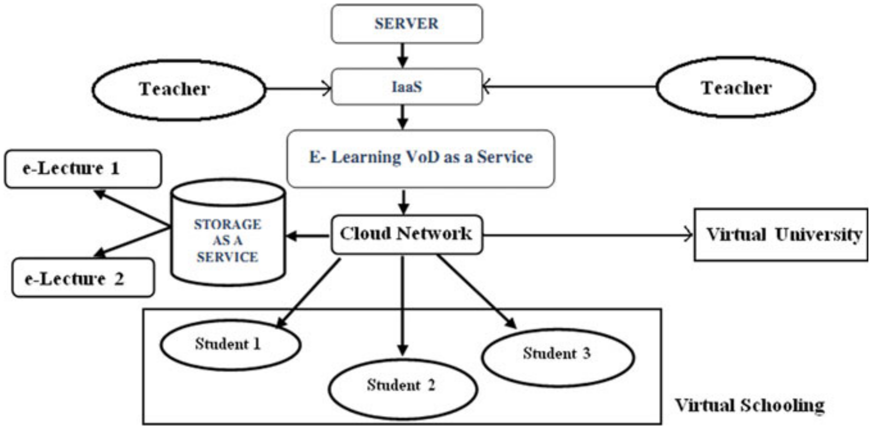


Fig. 10 Wireless cloud-based e-lecture operation module

free platform, has already a mature environment of Wireless Cloud services, and it has become the most popular online video community in global world. Therefore, it is decided to choose the virtual classroom technique as the Wireless Cloud computing service environment. And the e-learning will be integrated to the platform, and construct a robust module. Through a combination with Wireless Cloud computing, users can directly access the video from virtual classroom platform. It is not only reducing the cost of storage but also provide the assurance of service quality, such as bandwidth and video quality (Fig. 10).

9 Solutions for Mobility Applications in e-Education

Wireless Cloud computing solution can be used to enable seamless roaming between wireless networks to extend enterprise applications to Wireless Cloud workers. The capabilities provided by the Wireless Cloud IP solution create enhanced services for a variety of vertical applications. Vertical markets, such as trucking and transportation, healthcare, public safety and utilities, have realized the benefits that Wireless Cloud IP can offer to improve communications across the enterprise. The Wireless Cloud computing solution not only provides reliable wireless transport over wide-area networks (i.e., Wireless Data Network), but allows a Wireless Cloud data terminal to seamlessly migrate between a wireless wide-area network and a wireless LAN (Fig. 11).

This solution allows a Wireless Cloud data terminal to seamlessly migrate between a wireless LAN and a wireless wide-area network. The roaming features enable uninterrupted data service connections between the Server and a Wireless Cloud data terminal, which is also setup to relay data over wireless wide-area networks. While out of range of the wireless local area network, the Wireless Cloud employee,

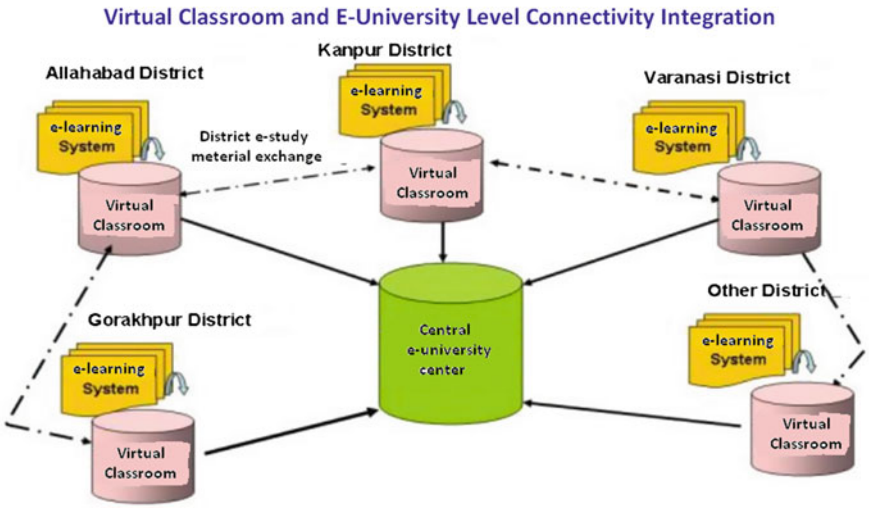
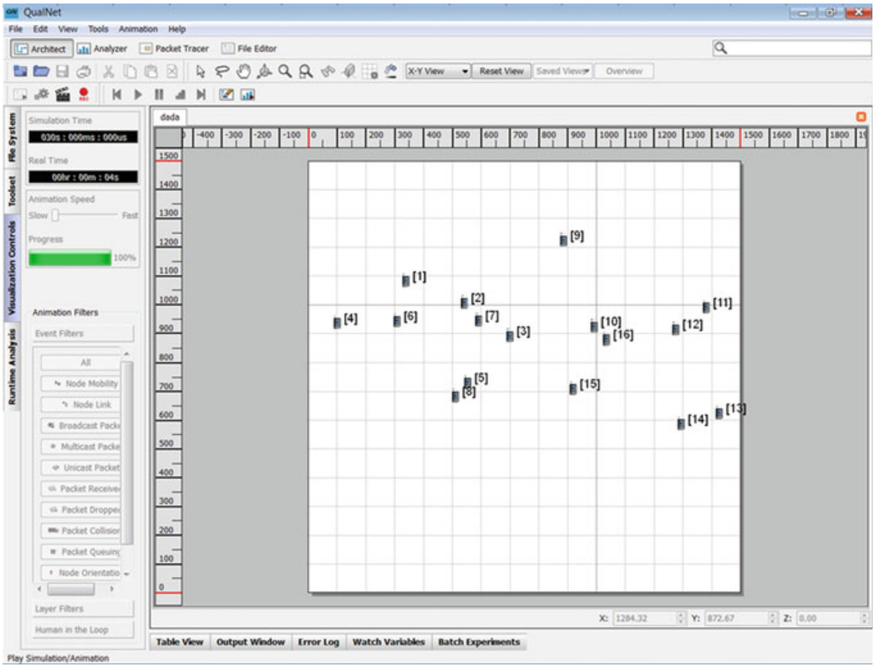

Virtual Classroom and E-University Level Connectivity Integration
Kanpur District
Allahabad District
Varanasi District
e-learning System
District e-study meterial exchange
Virtual Classroom
Central e-university center
Other District
Gorakhpur District

Fig. 11 Integration of e-university and country level virtual classrooms



Simulation Time 030s : 000ms : 000us

Real Time 00hr : 00m : 04s

X: 1284.32 Y: 872.67 Z: 0.00

Fig. 12 Development of virtual class room nodes in QualNet simulator

using a Wireless Cloud data terminal, has access to enterprise applications over a wireless wide-area network. However, as the Wireless Cloud data terminal comes into range of the fixed wireless LAN, the Wireless Cloud computing software automatically migrates to the fixed wireless LAN, allowing cost-effective access to enterprise information. In summary, the Wireless Cloud computing solution is intended to provide least-cost routing between a wireless LAN and a wireless wide area network.

10 Simulated Result in QualNet

The proposed could computing architecture based virtual classroom technique has been employed on QualNet software tool to simulate the tentative outcome from the e-education system (Fig. 12).

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled (Fig. 13).

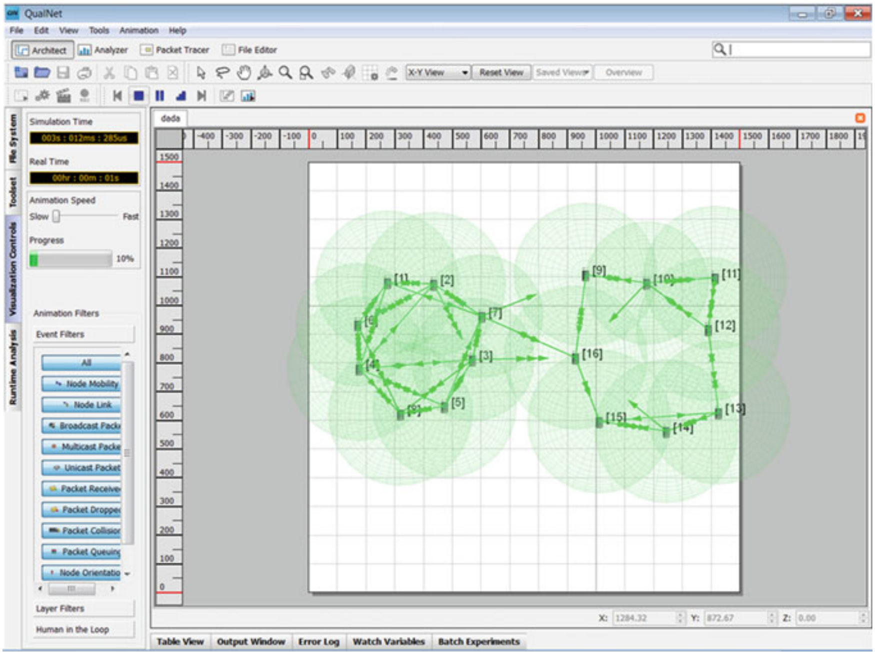


Fig. 13 Executing operational virtual class room nodes in QualNet

11 Conclusion

This paper presented an exploratory study from Allahabad district on how the use of a virtual classroom affects learning and teaching. A student Master class was for two sessions exposed to the use of the Centre, being observed during the sessions and interviewed afterwards. The activity theory is used to frame the study. This enabled to identify the changes in the activity system, and to analyse the effects on learning and teaching. The main finding is that the same learning outcomes as in the ordinary classroom may be achieved for students using virtual classroom technology under certain conditions. When considering the beginning of a virtual classroom as a full or partial substitute for an normal classroom, the findings suggest that the following major issues should be taken into account. Lecturer and students must accept that virtual learning sessions generally are more structured and predictable than learning sessions in the ordinary classroom. Students who already base their learning activity mainly on face-to-face interaction are likely to see the virtual classroom as an inferior substitute for 'the real thing'. There is a challenge to convince the students about the advantages of the virtual classroom and its relative learning importance. For communities of learners that are established with the virtual classroom as a primary medium of interaction from the start, conditions are extremely essential.

References

1. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) A survey on sensor networks. IEEE Commun Mag 40(8):102–114
2. Boonma P, Champrasert P, Suzuki J (2006) BiSNET: a biologically-inspired architecture for wireless sensor networks. In: Proceedings of the 2nd IEEE/IARIA international conference on autonomic and autonomous systems (IEEE/IARIA ICAS). Santa Clara, USA
3. Cheng Yu et al. (2006) A generic architecture for autonomic service and network management. Comput Commun 29(18):3691–3709. doi:10.1016/j.comcom.2006.06.017
4. Heinzelman W, Chandrakasan A, Balakrishnan H (2000) Energy-efficient communication protocol for wireless micro-sensor networks. In: Proceedings of the 33rd annual Hawaii international conference on system sciences, pp 3005–3014
5. Intanagonwiwat C, Govindan R, Estrin D, Heidemann J, Silva F (2003) Directed diffusion for wireless sensor networking. IEEE/ACM Trans Network 11:2–16
6. Kephart J, Chess D (2003) The vision of autonomic computing. IEEE Comput Mag 36(1):41–50

Unsupervised Methods on Image Database Using Cluster Mean Average Methods for Image Searching

R. Venkata Ramana Chary, K. V. N. Sunitha and D. Rajya Lakshmi

Abstract Image retrieval has been one of the most interesting and vivid research areas in the field of computer vision over the last decades. Image Retrieval systems are used in order to automatically index, search, retrieve, and surf image databases. Gathering of large collections of digital images has created the need for efficient and intelligent schemes for classifying and retrieval of images. In our proposed method, we are using Clustering Algorithm for retrieving the images from huge volumes of data with better performance. This requires image processing methods like color histogram feature extraction, classification of images, retrieval, and indexing steps in order to develop an efficient image retrieval system. In this work, processing is done through the image clustering method which is used for feature extraction taken place, classification is done using K-means [1] classification algorithm [2]. For retrieval of images, Euclidian distance method values are calculated between query image and database images. The main aim of this work is to extract images with similarity when the images are retrieved based on query image.

Keywords Color feature extraction · K-means · Mean values · Histogram · Threshold values

R. Venkata Ramana Chary (✉)
BVRIT, Narsapur Medak Dist, AP, India
e-mail: rvrchary@gmail.com

K. V. N. Sunitha
BVRIT Hyderabad College of Engineering for Women, Hyderabad, India
e-mail: k.v.n.sunitha@gmail.com

D. Rajya Lakshmi
GITAM Institute of Technology, Vishakapatnam, AP, India
e-mail: rdavuluri@yahoo.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_75,

1 Introduction

It is broadly understood that “A picture is worth a thousand words”. The meaning of an image is highly Individual and subjective. There is a fast growth in the size of digital image gathering. In many areas of government, academia, and hospitals, large collections of digital images are being created. Many of these collections are the product of digitizing existing collections of analog photographs, diagrams, drawings, paintings, and prints and a huge amount of information is out there. However, we cannot access to or make use of the information unless it is organized so as to allow an efficient browsing, searching, and retrieval. Retrieving an image from such large collections is a challenging problem, and thus methods for organizing a database of images and for efficient retrieval have become important. Content-based document image retrieval (CBIR), a technique, which uses visual contents to search images from large-scale image databases according to users’ interests, has been an active and fast advancing research area. During the past decade, remarkable progress has been made in both theoretical research and system development. Content-based image retrieval uses the visual contents of an image such as color, shape, texture, and decides the similarity between two images by reckoning the closeness of these different regions. The problems of image retrieval are becoming widely recognized, and the search for solutions is an increasingly active area for research and development.

Most of the approaches make use of the clustering algorithms for faster retrieval of images, though most of the time it does not come out with the most appealing results. Similarly, while retrieving images using shape-based methods there is a huge effort required to retrieve the images from the entire database.

In this paper we use clustering algorithms on image and collected features of the images to fiend the similar groups and similar images from the database we approached, mathematical methods and analyzed cluster mean values. All approaches are given good results from comparatively existing systems. The system performance, method implementations, and conclusions are discussed in the following sections.

1.1 Clustering Algorithms on Images

1.1.1 Clustering

Clustering or Cluster analysis is the task of allotting a set of objects into groups. Groups are called clusters when all objects in the same cluster are more similar than other groups objects. Clustering, a main explorative task in the specified domain is a commonly used technique for statistical data analysis in many fields, including image analysis, information retrieval, and pattern recognition.

Cluster analysis itself is not a specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with low distances among the cluster members, dense areas of the data space, intervals, or particular statistical distributions. Clustering can therefore be formulated as a multiobjective optimization problem. The appropriate clustering algorithm and parameter settings (including values such as the distance function to use, a density threshold, or the number of expected clusters) depend on the individual data set and intended use of the results. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multiobjective optimization that involves trial and failure. It will often be necessary to modify preprocessing and parameters until the result achieves the desired properties.

1.1.2 K-Means Clustering

K-means is the clustering algorithm used to determine the natural spectral grouping present in a data set [3]. This accepts from analysts the number of clusters to be located in the data. The algorithm then arbitrarily seeds or locates, that number of cluster centers in multidimensional measurement space [4]. Each pixel in the image is then assigned to the cluster whose arbitrary mean vector is the closest. We can calculate the mean of points in each segments of an image.
The K-means algorithm is an iterative method that is used to partition an image into K clusters [5]. The basic algorithm is:

- Pick *K* cluster centers, either randomly or based on some heuristic.
- Assign each pixel in the image to the cluster that minimizes the distance between the pixel and the cluster center.
- Re-compute the cluster centers by averaging all of the pixels in the cluster.
- Repeat the above second and third steps until convergence is attained.

2 Proposed Method

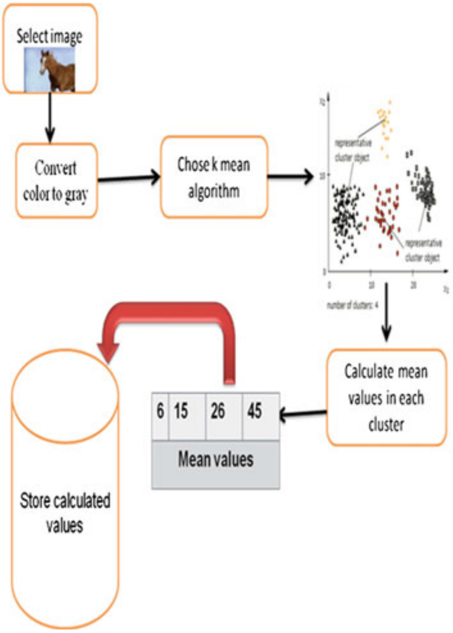
- Concentrated on visual contents of an image such as color and shape.
- Selected 10,000 image databases with common feature values.
- Extracted all image features separately.
- Proposed system implements features extraction methods used which are used for efficient retrieval.
- Extracted cluster mean values are detected and analyzed.
- This work provides a platform to extract images from the database using user query method.

2.1 System Architecture Implementation Process

From Table 1, we can understand the method and the implementation of the proposed system. In this architecture diagram, the user is asking the image as a query for searching similar images in the image database [6].

- In the first step, we have to find the threshold value of selected image and then search for it in a particular group which belongs to that threshold value.
- In the next step, color image gets converted into gray color and then k-mean clustering technique is applied.
- k value is selected in the form of 3, 4, 5, 6, 7, 8 depending on the user choice. From these user inputs, the system will compute the cluster mean values. Mean values are computed and stored into the database [7].
- For image similarity and comparisons, clustered mean values are used.
- Values are utilized from the database for different proposed methods.
- In the final step, similar images are produced by the proposed system.
- Final proposed methods are implemented and then we obtain the results.
- Table 1 shows the implementation process using clustering k-mean method with k values. Image is selected from the image directory, mean values and computed and stored in the database for image comparisons.

Table 1 System architecture diagram



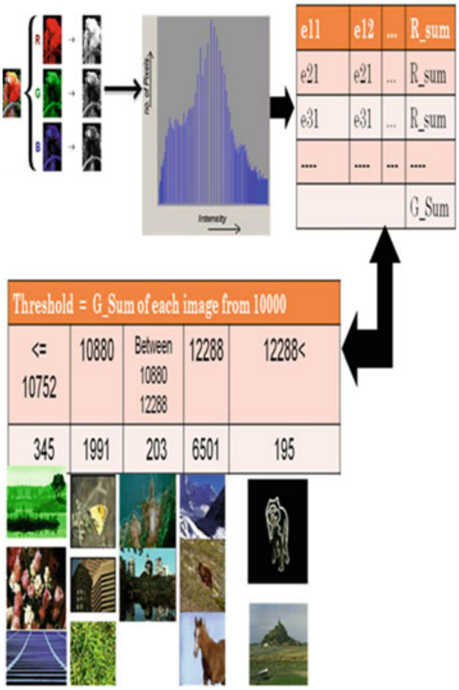
2.2 Grouping Images Using Threshold Values:

In this implementation process 10,000 images are selected so all images searching for similarity in sequential order means that it is very difficult to find the similarity of images. Instead of direct searching a new concept is implemented using calculation of the threshold values.

The process of the threshold calculation is presented in Table 2. In this process diagram explains the process implementation.

In the first step each image converts into the Gray format. From this image gray values are considered into the matrix format. Gray formatted image histogram values are taken into matrix form and row wise sum and grand row sums are calculated as shown in the diagram Table 2. This grand sum value is called it the threshold value. Using similar threshold values of all images are grouped. Using this method if the threshold values are matched then image is searching in that same group, otherwise no need to search in that group. In Table 4 we provided all possible groups with particular threshold values.

Table 2 Processing of threshold calculation



e11	e12	...	R_sum
e21	e21	...	R_sum
e31	e31	...	R_sum
....	....	...	....
			G_Sum

Threshold = G_Sum of each image from 10000				
<= 10752	10880	Between 10880 12288	12288	12288<
345	1991	203	6501	195

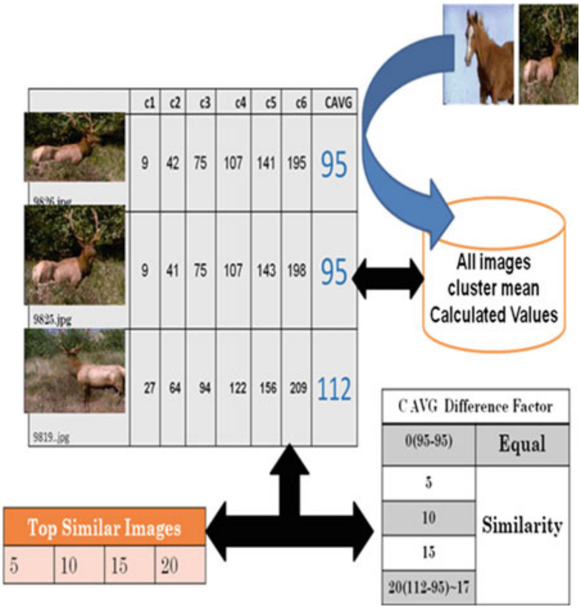
3 Propose Methods and Implementation

The proposed algorithm is summarized as follows:
Step 1: Input the query image into the process.
Step 2: Perform image threshold value based on gray image histogram and storing into database.
Step3: Select group of images in database based on the threshold values.
Step4: All selected images' cluster mean values are calculated using k-mean clustering method and all feature values are stored into image database.
Step5: The proposed method is implemented using the cluster mean values for searching.
Step6: Results and performance are evaluated using performance method.

3.1 Cluster Mean Average Method

After making the clustering on image cluster mean values are obtaining using k-mean method and all values are stored into the database file. From user mean difference factor is allowed in terms of equal to 0, not equality and similarity. Values start step-wise like 5, 10, 15, 20, 25 rage or top 5, 10, 15, 20, 30 images. Based on this images are obtained. This process is explained in Table 3. 9826.jpg, 9825.jpg are two

Table 3 Process of cluster mean method

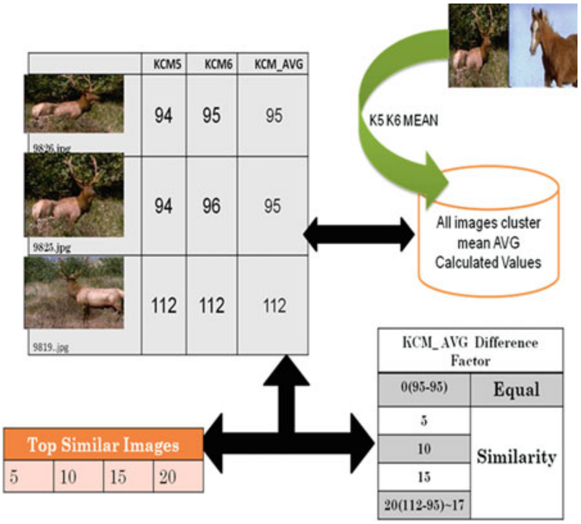


	c1	c2	c3	c4	c5	c6	CAVG
9826.jpg	9	42	75	107	141	195	95
9825.jpg	9	41	75	107	143	198	95
9819..jpg	27	64	94	122	156	209	112

C AVG Difference Factor	
0(95-95)	Equal
5	Similarity
10	
15	
20(112-95)~17	

Top Similar Images			
5	10	15	20

Table 4 Process of cluster mean of kc5 and kc6 average



	KCM5	KCM6	KCM_AVG
9826.jpg	94	95	95
9825.jpg	94	96	95
9819..jpg	112	112	112

KCM_AVG Difference Factor	
0(95-95)	Equal
5	Similarity
10	
15	
20(112-95)~17	

Top Similar Images			
5	10	15	20

images having the same CAVG value and both pictures are the same. At the same time 9819.jpg is similar and it is holed 112 CAVG value difference is only 7. Thus, we can conclude similar image groups holding deference values between 0 and 20 or near.

3.2 *Cluster Mean Average of kC5 and kC6 Method*

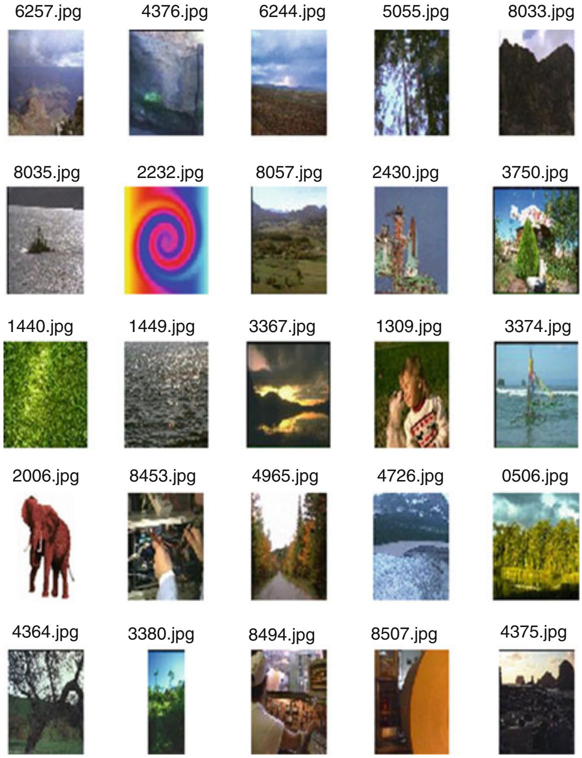
After making the clustering on image cluster mean values are obtained. All values are stored into the database file. Using the database file the mean value average for kc5 and kc6 are found and then average of kc5 and kc6 is made. Mean difference factor is allowed in terms equal to 0 when both images are equal, not equality and similarity. Values start step-wise like 5, 10, 15, 20, 25 rage or top 5, 10, 15, 20, 30 images. Based on this images are obtained.

4 Results Analysis

4.1 *Query Processing*

From the database table all images features are obtained using the MATLAB program, then the querying processing is implemented .

Table 5 Select fname, mean6, mavg from Kmeanx where mean6 between 125 and 135 ORDER BY mean6 ASC



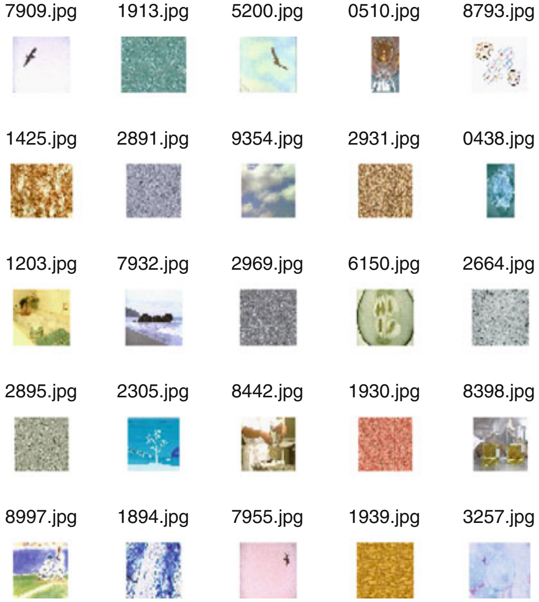
Example 1: query applied based on cluster mean 6 with difference 10 (125–130). Obtained result top 25 is shown in Table 5. From this table we can observe 17 images very similar to images retrieved from 10000 images.
Example 2: query applied based on cluster mean 6 with difference 5 (150–155). Obtained result top 25 is shown in Table 6. From this table we can observe very similar images obtained.

4.2 Method Performance Analysis

Proposed methods are implemented on different image groups. The proposed methods kC5, kC6, Kavg produced good results. The results are verified in top5, top10, top15 top25…top60.

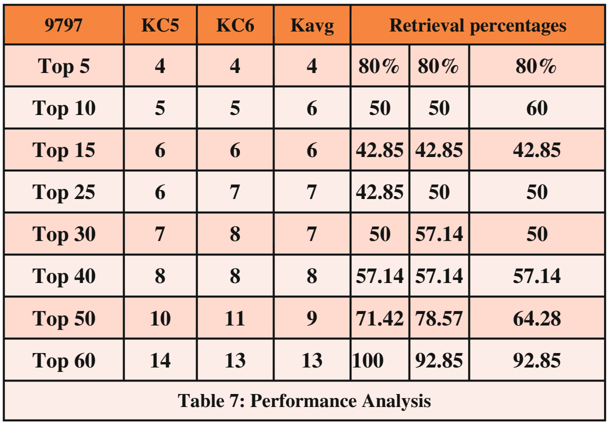
9797.jpg is taken as a query image and query is executed in that we observed that in **top 5 images four images got similar it is good** achievement in this verification all

Table 6 Select fname, mean6, mavg from Kmeanx where mavg between 150 and 155 ORDER BY mavg ASC



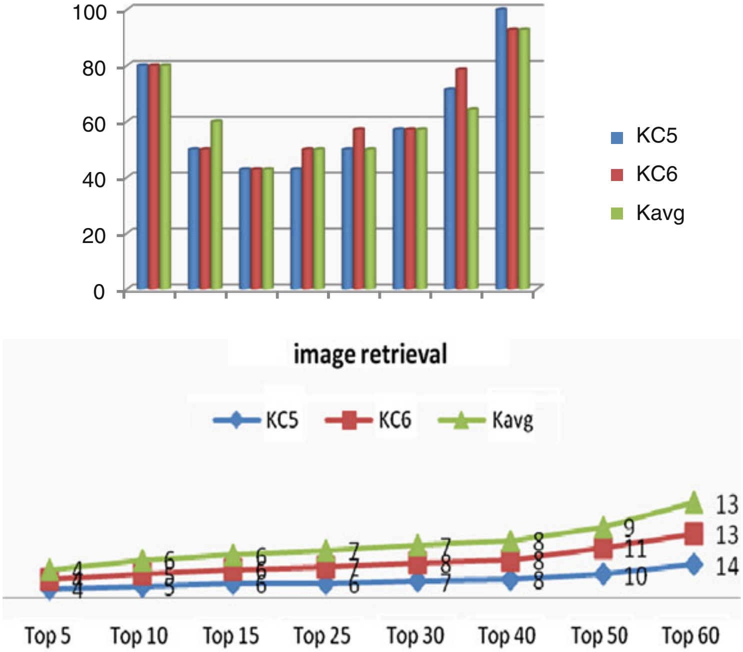
remaining results we can compare in Tables 7 and 8. Another important observation is kC5 to Kavg percent wise result is the same but retrieval time closer images are displayed adjacently.

Table 7 Performance analysis



9797	KC5	KC6	Kavg	Retrieval percentages		
Top 5	4	4	4	80%	80%	80%
Top 10	5	5	6	50	50	60
Top 15	6	6	6	42.85	42.85	42.85
Top 25	6	7	7	42.85	50	50
Top 30	7	8	7	50	57.14	50
Top 40	8	8	8	57.14	57.14	57.14
Top 50	10	11	9	71.42	78.57	64.28
Top 60	14	13	13	100	92.85	92.85
Table 7: Performance Analysis						

Table 8 Performance analysis



5 Conclusion

In this paper we presented an approach for Image Retrieval using cluster mean method using K-Means clustering techniques, where images are initially clustered into groups having similar threshold values. Image feature values are obtained from the images and stored into the database. Clustering assists faster image retrieval and also allows the search for most relevant images in large image databases. The proposed clustering method based on the optimization of an overall measure of clustering quality is known for its efficiency in producing accurate results in image retrieval. Along with that we further refine the retrieved set of images based on clustering method as there may be irrelevant images in the result set. Using the proposed methods we can retrieve very relevant images. This method has shown good performance.

References

1. Murthy VSVS et al (2010) Content based image retrieval using hierarchical and K-means. Int J Eng Sci Tech 3:209–212
2. Swain MJ, Ballard DH (1991) Color indexing. Int J Comput Vis 7(1):11–32

3. Tonge VG, Content based image retrieval by K-Means clustering algorithm. Int J Eng Sci Tech 2:209–212 (February 2011 NCICT conference special issue)
4. Maini R, Aggarwal H (2009) Study and comparison of various image edge detection techniques international. J Image Process 3(1):1–12
5. Striker M, Orengo M (1995) Similarity of color images. Proc SPIE Storage Retrieval Image Video Database 2420:381–392
6. Sunitha KVN (2012) Feature extraction methods for color image similarity. Adv Comput Int J 3(2): 2229–6727 [Online], 2229–726X [Print]
7. Venkata Ramana Chary R, Rajya Lakshmi D, Sunitha KVN (2012) Image retrieval techniques for color based images from large set of database. Int J Comput Appl Found Comput Sci 40(4). ISBN: 978-93-80866-44-11

Impact of Fix Cluster Head Selection (FCHS) Routing Protocol for Wireless Sensors Network

Priyanka Chugh Shivanka and Ashwani Kumar

Abstract Today main objective of wsn is to minimize the energy dissipation for the whole network. Clustering is one of the most knows methods widely used to face these challenges. Fix Cluster Head Selection (FCHS) routing protocol analyses how the system lifetime is improved by fixing the selection of clusters head. In this paper a cluster based communication protocol with considering the low energy consumption in wireless sensor networks, is introduced which balanced the energy load among the sensor node. Fix Cluster Head Selection (FCHS) analyses the cluster head selection to find the optimal probability of becoming a cluster head. Simulation result of FCHS corresponding to LEACH in mat lab shows that these designs increase the lifetime of the network 105 %. We found that FCHS yield longer stability region for higher values of extra energy brought by more powerful nodes. Finally, LEACH protocol and the improved algorithm simulate in MATLAB, and make Performance analysis and comparison in number of nodes alive, total energy consumption in network and a round of node death distribution. The results show that, FCHS compared with LEACH protocol, the improved algorithm prolongs the network life cycle, raises energy utilization and has good load balance.

Keywords Fix cluster head selection FCHS · Wireless sensor network · LEACH · Energy dissipation · Routing protocol

P. C. Shivanka (✉)
CSE, MMU University, Ambala, India
e-mail: shivankachugh84@gmail.com

A. Kumar
EE, Lingaya's University, Faridabad, India
e-mail: ashwani_smash@yahoo.co.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_76,

1 Introduction

A sensor network is composed of a large no of sensor nodes, which are densely deployed either inside the phenomenon or very close to it. The no of nodes in sensor network can be several orders of magnitude higher than the nodes in an ad hoc network. Sensor network are densely deployed [9]. This tiny system, which consist of sensing, data processing and communication components are known as wireless sensor network [6].Sensor nodes are limited to power, computational capacities and memory. Sensor nodes mainly use broadcast, most ad hoc network are based on Point to Point. These nodes, with limited computing, communicating, and sensing capabilities as well as Limited energy can make the best use of themselves together data from sensor node to Base Station (BS) by using excellent network topologies, optimized routing scheme [8]. In simplest method these node get data from sensor to BS by using topology, the energy loss in LEACH is more than the Fix Cluster Head Selection (FCHS). WSN has limited supply of energy hence an energy conserving routing protocol is important for these networks. Therefore, while conventional network aim to achieve high quality of service provisions, sensor network focus primarily on power consumption. There are several power-aware routing protocol explained in the literature [2, 7, 8, 10]. Although the clustering can reduce the energy consumption, the main problem is that the energy consumption is concentrated on the cluster heads. In order to overcome this demerit, the issue in cluster routing is how to distribute the energy consumption. The representative solution is LEACH [3] which is a localized clustering method based on the probability model. All sensor nodes evenly elect itself as a cluster head based on the probability model to distribute the energy consumption. However, in some cases, in-efficient cluster heads can be elected. Because LEACH is only depend on probability model, some cluster heads may be very close each other and can be located in the edge of the WSNs. These in-efficient cluster heads could not maximize the energy efficiency [4]. Recently, Gupta [4] introduced a cluster head election method using fuzzy logic to overcome the defects of LEACH.

2 FCHS Algorithm Details

The FCHS was created with particular criteria such as Fixed cluster head selection, fixed location of sensor node that will cause less energy dissipation of cluster head and base station so that the protocol becomes computational simple as the static route could be used. FCHS is source initiates protocol with time driven reporting, so the sensor node would always have data to send to the base station. FCHS would also apply data aggregation to avoid information overload or access of data, in this aspect FCHS shares the same features as LEACH [5]. FCHS shows an analytic architecture for obtaining the stable probability with which a node would become a cluster head in order to minimize the network's energy consumption. The analysis presented for a small network. FCHS change the choice of cluster head over time for minimum energy dissipation between the threshold level 0 and 1 to become clusters head (Table 1).

2.1 Proposed Fchs Algorithm for Cluster Head

Step1: Initialization
Cluster head change over time for minimum energy dissipation Decision is made by random number between 0 and 1 node become Cluster Head.
p = percentage of cluster head.
r = current round.

Step2: Header Selection

- If the number is less than the following threshold:
- if(temp_rand< =(p/(1- p*mod(r,round(1/p))))
- if(countCHs< =5)
- countCHs = countCHs + 1;

Step 3: Advertisement

- After the selection of Cluster Head choose in between the threshold
- The node broadcast a CHs_ADV advertisement to normal node.
- Formation and function of Cluster head.
- Next round is executed until rmax value.

Step 4: Management

- Member nodes of each cluster send data to CHs.
- CHs collects the data
- CHs send the collected data to the BS.

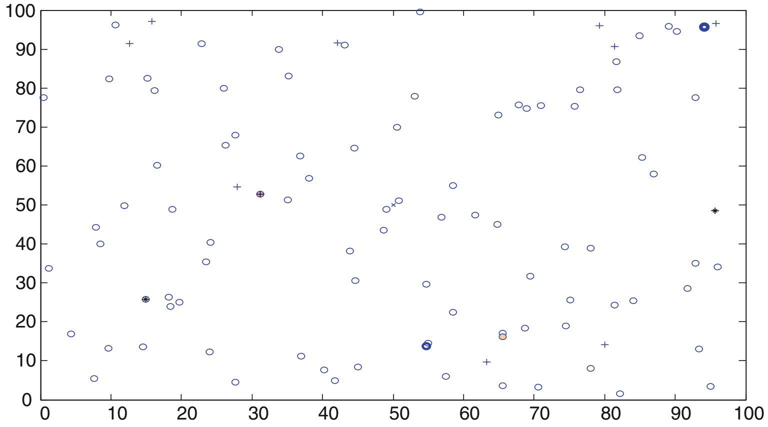
Step5: Header Switch

- If E(CH)<E', the node become a header
- If E of the node greater than E', go to step2 total energy of the network
- Remaining_Energy = 0;
- Total_Energy = n*((1-m)*Eo+m*(1+a)*Eo);
- for i = 1:n
- Remaining_Energy = Remaining_Energy + S(i).E;
- end
- Energy_Consumed = Total_Energy-Remaining_Energy;

Table 1 System lifetime using different amounts of initial energy for the sensors

Initial energy (J/node)	Protocol	Probability of Cluster head	Rounds first node dies	Round last node dies
0.1	FCHS	CH = 4	105	2113
		CH = 5	135	3113
		CH = 7	167	1576
		CH = 10	184	904
	LEACH	CH = 4	96	612
		CH = 5	123	637
		CH = 7	180	579
		CH = 10	177	737
0.25	FCHS	CH = 4	412	2005
		CH = 5	462	3430
		CH = 7	485	2085
		CH = 10	495	2571
	LEACH	CH = 4	275	2071
		CH = 5	353	1295
		CH = 7	413	1395
		CH = 10	492	2549
0.5	FCHS	CH = 4	541	2768
		CH = 5	645	3583
		CH = 7	740	3958
		CH = 10	853	3972
	LEACH	CH = 4	603	2846
		CH = 5	595	3000
		CH = 7	835	2590
		CH = 10	948	3000
1	FCHS	CH = 4	1681	10000
		CH = 5	1730	10000
		CH = 7	1932	8621
		CH = 10	2019	4643
	LEACH	CH = 4	1262	4197
		CH = 5	1200	4014
		CH = 7	1536	7000
		CH = 10	1923	9000

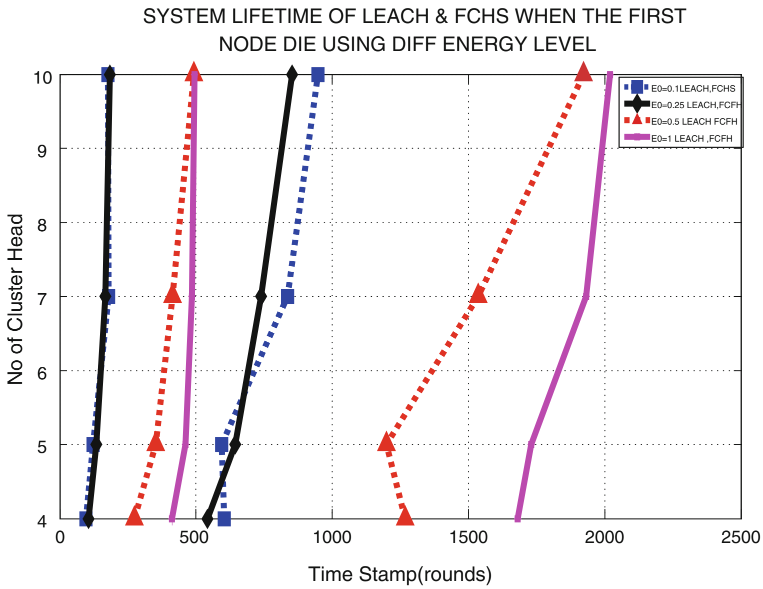
2.2 Simulation Results for FCHS and LEACH



3 Comparison of FCHS and LEACH When the First Node Dies

Performance graph of System lifetime when the *First node die* using Different Energy Level with fixing no of cluster head shows the comparison of the FCHS and LEACH Initially, when clusters are being created, each node decides whether or not to become a cluster-head based on the suggested percentage of cluster heads for the network (determined a priori) and here in this graph Dotted (---) line shows when the first node of LEACH gets die shows the life time of the network system the number of times the node has been a cluster-head so far. This decision is made by the node n choosing a random number between 0 and 1. If the number is less than a threshold T(n)[1], the node becomes a cluster-head for the current round. The threshold is set as:

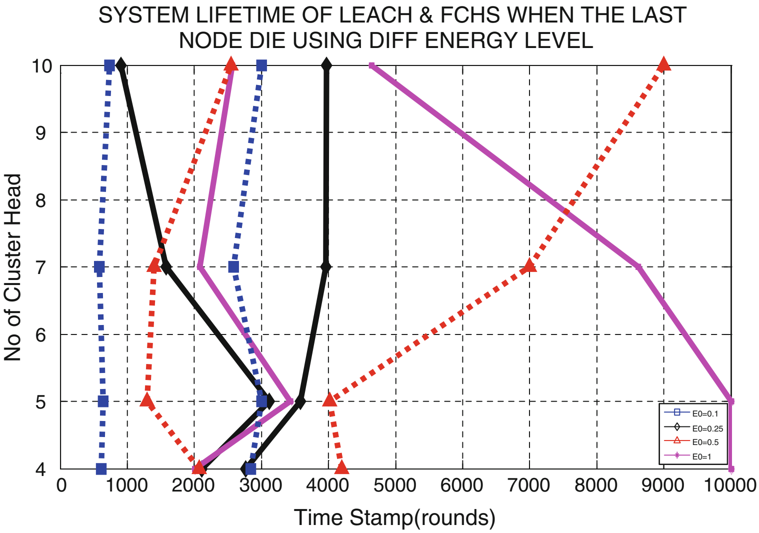
$$T(n) = \begin{cases} \dfrac{p}{1 - P * (rmod\frac{1}{P})} & \text{if } n \in G \\ 0 & \text{otherwise} \end{cases}$$



Where P = the desired percentage of cluster heads (e.g., P = 0.10), r = the current round, and G is the set of nodes that have not been cluster-heads in the last 1/P rounds. Using this threshold, each node will be a cluster-head at some point within 1/P rounds. During round 0 (r = 0), each node has a Clustering and Cluster-Head Selection Techniques used in WSN probability P of becoming a cluster-head. In the graph the *rigid line* shows the lifetime of the cluster head in the FCHS. Hence, the FCHS *increase the lifetime of the network* by fixing the choice of cluster head.

3.1 Comparison of FCHS and LEACH When the Last Node Dies

Performance graph of System lifetime when the *Last node die* using Different Energy Level with fixing no of cluster head shows the comparison of the FCHS and LEACH Here in this graph Dotted (----) line shows when the Last node of LEACH gets die shows the life time of the network system the number of times the node has been a cluster-head so far. In the graph the rigid line shows the lifetime of the cluster head in the FCHS. Hence, the FCHS increase the lifetime of the network by fixing the choice of cluster head by increase the value by 1000 rmax [4].



4 Conclusions

In WSN power efficiency is an important performance metrics that must be considered when designing the routing protocol. Objective of this paper is to design an energy aware routing protocol based on LEACH. Wireless Sensor networks (WSN) have emerged as a promising tool for monitoring (and possibly actuating) the physical world, we also compared the routing protocols according to different parameters since one single Routing protocol is not applicable in all the situations. We have performed detailed analysis of the LEACH protocol and the Direct Communication Protocol by running the simulations in MATLAB and by fixing the no of cluster head it enhanced the System Life time also the less energy dissipation. We have found that the LEACH protocol outperforms the Direct Communication Protocol in terms of Energy Dissipation. To form constant number of clusters Results show that FCHS is an efficient routing protocol when compared to conventional routing protocol LEACH, which has shown less lifespan of network. Simulation results show that FCHS reduces communication energy by as much as 105 % compared to LEACH. This is due to the fix cluster head selection, which results in optimal probability of becoming a cluster head. It could be concluded that an effective selection of a cluster head could reduce the usage of consumption power. Simulation also shows by minimizing the transmission distance of FCHS and LEACH would result in increase the lifetime of the network.

References

1. Akkaya K, Younis M (2005) A survey on routing protocol for wireless sensor networks. Elsevier Ad Hoc J 3(3):325–349
2. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) A survey on wireless sensor networks. IEEE Commun Mag 40(8):102–114
3. Chandrakasan A, Balakrishnan H (2000) Energy efficient communication protocol for wireless sensor networks. In: Proceedings of the 33rd hawaii international conference on system science, January 2000
4. Gupta I, Riordan D, Sampalli S (2005) Cluster-head election using fuzzy logic for wireless sensor networks. In: Communication networks and services research conference, Halifax, 255–260 May 2005
5. Ibriq J, Mahgoub I (2004) Cluster based routing in wireless sensor networks:issues and challenges. In: Proceedings of the 2004 symposium on performance evaluation of computer telecommunications System, USA, 759–766 July 2004
6. Ilays M (2004) Imam Mahgoub, handbook of sensor networks: compact wireless and wired sensing systems. CRC Press, Boca Raton, USA
7. Jiang Q, Manivann D (2004) Routing protocols for sensor networks. In: Proceedings of the 1st IEEE consumer communications and networking conference, Las Vegas, USA, 93–98 January 2004
8. Kim HS, Han KJ (2005) A power efficient routing protocol based on balanced tree in the WSN. In: Proceedings of the 1st international conference on distributed framework for multimedia applications, 138–143 February 2005
9. Norsyafizan W, Muhamad W, Dimyati K, Mohamad R (March 2010) Abd Kadir R, Heinzelman WR (2010) Evaluation of stable cluster head election (SCHE). In: Routing protocol for wireless sensor networks
10. Selvakennedy S (2006) An energy efficient clustering algorithm for multihop data gathering in wireless sensor networks. J Comput 1(1):40–47

Part X
Workshops: The Fourth International Workshop on Ad Hoc and Ubiquitous Computing (AUC- 2012)

Ontology Oriented Approach to Service Selection and Invocation in Complex Context Analysis

Slawomir Nasiadka

Abstract Context-aware applications running in the intelligent space are taken into account and their execution in the service oriented environment is considered. It has been presented where and how SOA services can be utilized during their execution: to analyze current context of the application and to support execution of strictly determined actions suitable for that context. The proposed mechanism of context-aware service selection and invocation is based on two items: ontology description of the context and a model of available services. It was shown that proposed solution is less efficient but more flexible than direct context analysis.

1 Introduction

Context-Aware Applications (CAA) [5] are applications that are executed in the Intelligent Space (IS) [6], adapt to its changes and support its users. Their definition consists of a set of pairs, each of which includes expected conditions of an action invocation and adaptation action. Expected conditions are checked against current context (created from context data delivered by sensors) and if the context meets conditions then adaptation action is invoked. The context analysis performed during checking of conditions can be complex (e.g. comparing current context with previous one) and as such it may be enhanced with the usage of SOA services (called supporting services) which makes it SOA oriented [4]. They can be selected, matched and invoked dynamically which is crucial for applications running in the IS, as it is a highly dynamic environment. This approach enables context analysis mechanism to adapt to current situation and capabilities (in the sense of available services) of the IS. The aim of the article is to propose such an adaptable mechanism. To be able to perform that enhanced analysis there has to be created a model of a service.

S. Nasiadka (✉)
Gdansk University of Technology, Gdansk, Poland
e-mail: slawomir.nasiadka@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_77,

It is presented how to combine that model with the IS description to achieve proper services selection and automatic invocation during context data analysis. Proposed mechanism has been implemented and evaluated. That allowed to compare its time characteristic with the approach that does not use services during context analysis.

The paper is organized as follows. Section 2 contains state of the art regarding services matching. In Sect. 3 there is described usage on context in CAA application and a sample application is presented. Section 4 enumerates types of services used during CAA execution and introduces a proposition of mechanism for searching, matching and invoking SOA services. Section 5 describes its evaluation based on the implementation of sample application and presents experimental results. In Sect. 6 there are discussed conclusions and future work.

2 State of the Art

In recent years substantial efforts have been made for designing different algorithms for services matching. In general they can be divided in two categories: syntactic and semantic service matching methods. Syntactic methods (like JINI, UDDI, UPNP) directly compare strings describing services and their interfaces. Their main drawback is lack of precision in service discovery and matching process which results in low efficiency of that approaches. This is because the same functionality or interface included in a service can be described using different concepts, however having the same meaning. In opposite, describing concepts may be equal but may have different meaning. Systems falling into the second category utilize semantic approach. Great number of them, like [2, 9, 10], mostly concentrate on computing a level of similarity between services based on their description and interfaces. This is so called functionality-based semantic service matching approach, where the semantic of services is captured through semantic description of inputs, outputs, preconditions and effects (IOPE) [7]. Semantic interpretation of the description of service functionality and interfaces is realized with a help of some sort of knowledge, expressed e.g. using ontology. For instance in [11] there has been proposed an approach for efficient service matching algorithm using query rewriting based on ontology. Other works focus on additional parameters that are important during services matching commonly called QoS attributes (like trustworthiness, speed of execution and so on) [1, 8]. However, those approaches mostly concentrate on the service interfaces or messages themselves, lacking analysis and comparison of complex data types that can be exchanged between services (as parts of messages). Such data types are commonly used in CAA, where context that is used by applications is complex. In [3] authors propose Web Service discovery method based on schema matching, which regards comparing data types that may be complex. However, it doesn't use semantics to describe structure of the data type in a way that would allow to check whether two types are equal (and in effect could be transformed automatically during service method invocation).

3 Context in CAA

Expected conditions of an action invocation defined in CAA can be complex. This may lead to express them in a programming language (like Java or C#). That further makes them hard to maintain and limits adaptability to the current IS capabilities (available services). Solution for that issues is to combine a notation based on rules with a SOA approach for expressing expected conditions and for checking whether they are met by the current context. Rules represent high level description of conditions which consists of semantic description of what services have to be used (e.g. checking whether it is rush hour time) and how they are connected (e.g. logical relation and). Services can be then invoked during checking of a particular condition.

As a sample CAA it will be used an application that monitors traffic and automatically charges drivers for using the road. It uses cameras which allow to identify passing cars. However, using the road costs more during rush hours than other times of the day (it is assumed that charge is a logarithmic function of the traffic volume). Rush hours for the particular road can appear during different parts of day, can change from day to day, they are different for different parts of the road and depend on some external factors like the closure of roads in the neighborhood (e.g. caused by road works). Depending on the weather (level of lightness) the cameras should use different lights to achieve the most clear picture of the car and its numbers. That adaptation should depend not only on the current level of lightness but also on the weather forecast so that the application could adjust the level of camera lights more precisely. Similarly, for great traffic volume the application should use faster (and maybe less precise) algorithms of processing the video stream, whereas when the volume is low the stream analysis can be done slower and more precise. Those algorithms should be chosen in advance based on the traffic patterns from road usage history. There has been made an assumption that within the IS there are sensors delivering value of current time, current level of lightness and current traffic volume. However, they are only partial data needed for recognition whether any of the actions should be invoked and more advanced analysis is necessarily. The first step needed for that analysis is to unify the context data representation.

Sensors that are located in the IS are heterogeneous by nature and can be treated as services (in the sense of SOA) that deliver data. They may have different capabilities and may represent the same value in different ways (e.g. one sensor can represent volume of traffic as a property of the road and another the other way round). In other words each sensor can use different ontology for describing data it delivers. However, the process of checking whether context meets expected conditions of an action invocation has to understand that data in a uniform way. Hence, all data delivered from sensors should be transformed into a standard representation. As the IS is a complex construct, that contains objects, parameters and relations between them, a good choice for creating such representation is the ontology.

The ontology that has been constructed to describe the sample CAA, IS and services running in it is presented in Fig. 1.

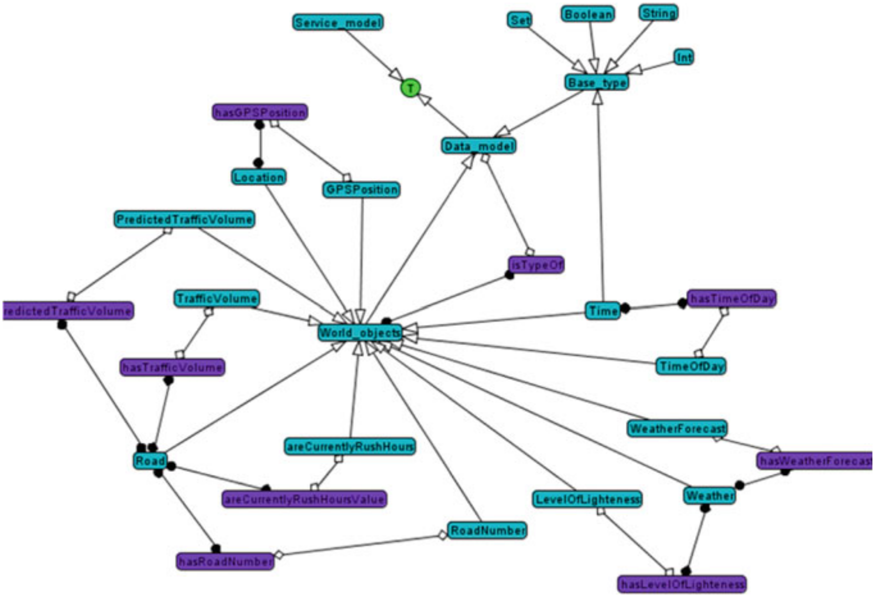


Fig. 1 Top level view of the ontology for sample CAA application

It consists of two main classes: *Data_model* and *Service_model*. First one represents all the concepts that can be used to describe the IS. They are further divided into two categories: *World_objects* and *Base_type*. *World_objects* represent objects that exist in the IS (like road, weather). *Base_type* represent types of values that can be assigned to parameters associated with IS objects (like string, int). The *isTypeOf* property allows to express that parameters of objects are a type of a class that inherits from *Data_model* (e.g. name (parameter) of the car (object) is a type of string). Both context and expected conditions of an action invocation can be expressed using that representation. It is also crucial for automatic selection and invocation of services during execution of CAA.

4 Services Usage During CAA Execution

Services used during CAA execution can be divided into two categories. First one represents supporting services used context analysis and the second contains services used for action execution.

4.1 Types of Services

During checking whether expected conditions of an action invocation are met, it may be necessary to perform complex operations. Moreover, additional requirements like reliability, quality (QoS) and privacy may need to be taken into account. That is why the process of context analysis should be based on the usage of supporting services. During execution of the sample application, they are utilized for e.g. determining the weather based on the level of lightness and the weather forecast.

SOA services are also used during actions execution, which can be described as a scenario (expressed in e.g. BPEL). Adaptation actions defined in the sample application include e.g. using faster/slower algorithms for video stream processing. CAA execution consists of the usage of presented types of services in three different areas: context data delivery, context data analysis and actions execution. This article focuses on the second aspect. To be able to find a service in a process of checking whether to invoke an action there has to be created a model of a service.

4.2 Service Model

To create a model of a service it is necessary to define basic concepts and relations between them. The model has to be able to express what a service can be used for, what are its methods, how to invoke them and what are their preconditions. It should also allow to automatically invoke a method of the service (with passing context data). Proposed model is presented in Fig. 2.

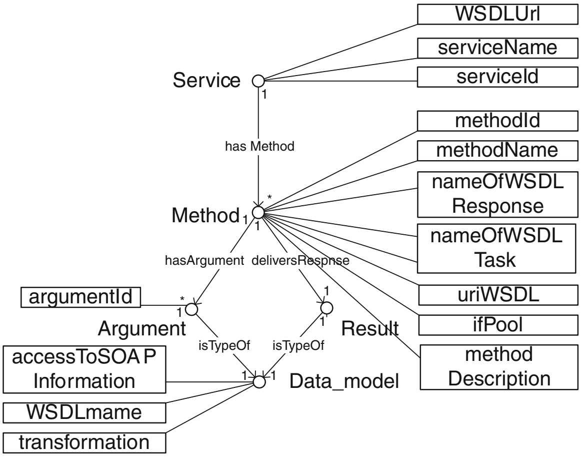


Fig. 2 Model of a service

Concepts used in the proposed model are included in the ontology presented in Fig. 1 and inherit from *Service_model* concept. There has been assumed that Web Services are representative for services in general (in the sense of SOA). The model, allows to enrich process of searching for a service with recognition if its methods can be invoked on a particular set of context data (for example whether service that checks if it is a rush hour time can work on time of the day that is represented as a string value). Each service has some global *id* and global *name*. For automatic service invocation the CAA needs to know low level syntactic description of a service (called grounding in OWLS). That is why a model has a pointer to that description (*WSDLUrl*). Further, each service has one or more methods which also have their ids (*methodId*) and names (*methodName*). For communication based on Web Service standard there is also a namespace needed which is contained in the *uriWSDL* attribute. A *description* of a method allows for its semantic search and attribute *ifPool* allows to express whether the method will deliver a context data on its own or it needs to be pooled. The last two attributes of a method (*nameOfWSDLTask* and *nameOfWSDLResponse*) are used to construct messages exchanged with a service. Each method has one result and have zero or more arguments (each of which is identified by its own *argumentId*) which are described using concepts that inherit from the *Data_model*. Attributes *WSDLname*, *accessToSOAPInformation* and *transformation* are used for manipulation (based on XSLT) of a SOAP message (needed for appropriate context data transformation to and from uniform representation—see Sect. 3).

A description of a service (supplied by a service provider) is added to the ontology in a form of individuals (of a service, a method, an argument and so on). Data that each method works on is also represented in a form individuals. In case of complex data type, the representation in the ontology includes its structure. For instance, the ontology contains classes *Road* and *TrafficVolume* which are connected with a property *hasTrafficVolume*. Additionally, *TrafficVolume* is connected with a class *Int* using relation *isTypeOf*. The description of the service method working on the context data of traffic volume would include individuals of classes *Road*, *TrafficVolume* and *Int* connected with properties *hasTrafficVolume* and *isTypeOf*. Values of *WSDLname*, *accessToSOAPInformation* and *transformations* included in the description would then contain information on how to create a SOAP message with information about road that has to be passed to the service.

4.3 Service Selection and Invocation Mechanisms

During searching for a service, the process of context analysis has to indicate what function should the service (and its methods) perform and which context data it can use. Hence, the first step is to find a service (and its method) based on the semantic description contained in the expected conditions of an action invocation (an *methodDescription* attribute is used for that purpose). Next, there should be performed a check whether the method can be invoked using currently available context data. Expected conditions of an action invocation include information about which types of data they operate on (for example a method that checks if it is rush hour time utilizes the number of the road, time of the day and GPS position). That creates a specification of the interface that the service method should support. Because both conditions and service methods descriptions are expressed using concepts from the same ontology, the selection process compares parts of that ontology. That process is organized as follows:

- find out what is the class of the individual in the compared expected condition context data and in the method: if they are the same—go to the next step; if no—try next method,
- for each property contained in the individual (from the expected condition context data) check if the individual that models data used by the service method has the same property,
  - if no, then try next method,
  - if yes, perform recursive check (starting from first bullet) for the individual that the property (from the expected conditions) points to; if there are no further properties in the expected conditions, check if there are none in the method as well. If there are, try next method; if there aren't, the method argument is suitable.

When all arguments are suitable (have been successfully matched with the context data) then the method can be used. Invocation is done by constructing a SOAP message. The header of the message and the envelope is created using attributes from the model of service (which includes a method description). The body is created using transformation defined in the model which allows to prepare an appropriate XML document using the ontology description of the IS. Thanks to that the service itself can use different ontology for describing data passed to and from its methods.

Presented approach doesn't make any assumptions about data structure as well as how the data is exchanged with a service. Because of that, it supports any kind of data that can be described using WSDL language. It is important that the proposed mechanism needs to achieve a perfect match between required description contained in expected conditions and service description. This is because not a complete match wouldn't allow to automatically invoke a service.

5 Evaluation of Mechanism

The approach described in this article has been implemented within execution environment for CAA applications. All SOA services have been created as Web Services in .NET technology and were running within the IIS server. For operations on the ontology there has been used Jena framework and SPARQL language. Example of definition of expected conditions of an action invocation from the sample CAA application is presented in Listing 1. It regards checking if it is rush hour time based on the number of the road, time of the day and GPS position.

Listing 1 Definition of an expected conditions of an action invocation for sample CAA application

```
<ContextDefinition>
  <Variables>
   <Variable name="Road" def="Ont:Road"/>
   <Variable name="Position" def="Ont:Location"/>
   <Variable name="Time" def="Ont:Time"/>
  </Variables>
  <Relations>
    <Relation name="rel1" expr="isRushHour([var:Road;
              //Ont:roadNumber],[var:Location;
              //Ont:GPSPosition])"
              groups="Road, Location" />
    <Relation name="rel2" expr="isEqual([var:Time;
              //Ont:timeOfDay],
              {Ont:StringGlobal;"day"})" groups="" />
    <Relation name="mainRel"
                expr="and([rel:rel1],[rel:rel2])"
                main="true" groups="" />
  </Relations>
</ContextDefinition>
```

The definition contains a list of variables and relations that operate on those variables. Variables are set (filled) when the context data (delivered by the sensors) that has a type pointed in the *def* attribute of the variable. In the example there are three relations: *rel1*, *rel2* and *mainRel*. First one invokes external Web Service that checks if it is rush hour time based on the number of the road and the GPS location. The *group* attribute allows to express that both values should be delivered by the same sensor (to avoid the situation that the road number and GPS location are delivered by two different sensors located on different roads). Method description that is searched for in the ontology is *isRushHour* and the method interface needs to accept *roadNumber* and *GPSLocation* (types are defined in the ontology). In case of a second relation the time of day should be compared with the string value. In the ontology there has been defined a type named *StringGlobal* that was used in the definition of the relation. In effect the search is made for a method that has method description of *isEqual* and accepts parameters of types *timeOfDay* and *StringGlobal*. The last relation (*mainRel*) performs logical function *and*, that returns *true* when both *rel1* and *rel2* are *true*.

Based on the prepared implementation there has been made a research on how fast the CAA application invokes an adaptation action as the response to a new context. Definition of the application contained a single expected condition of an action invocation (see Listing 1) and an action of setting higher charges during rush hours. One supporting service analyzed the rush hour patterns for a part of the road based the road number and GPS location of the sensor that observed the road, and another checked time of the day. The research was made for different sizes of context

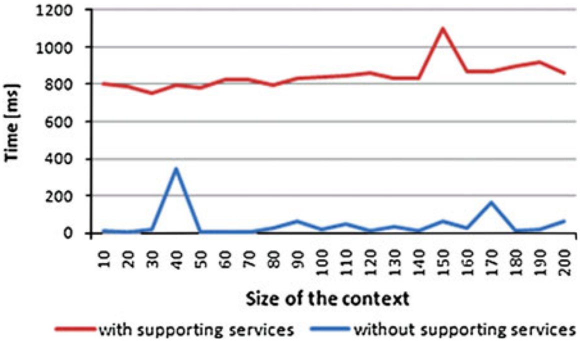


Fig. 3 Time of conditions check depending on the size of the context

which was counted as the number of parameters that are used in the conditions (in the example from Listing 1 size of the context is 3). That size had values ranging from 10 to 200 with a step of 10. To be able to make such an experiment there has been assumed that the road can only be identified by a set of numbers each of which can be described using different (single) object from the ontology that is associated with a single parameter. For comparison the experiment was repeated for the same application definition, but instead of invocation of supporting services all processing has been done by the context analysis process itself. The results have been presented in Fig. 3.

As it can be seen, values of time needed to analyze context data continuously increase along with the increasing size of the context. In case of analyzing context data without supporting services the growth in the time is very small. It can be observed that in the sample implementation the process of checking whether expected conditions of an action invocation have been met is much slower when supporting services are used.

6 Conclusions and Future Work

The mechanism of service oriented context analysis proposed in this paper allows to search, select and invoke supporting services during context analysis. The selection utilizes ontology to semantically compare complex data structures to check whether a service has compatible interfaces, and the invocation is enhanced with semantic data transformations that ensure proper data adaptation for service method execution. Usage of this mechanism causes slower performance than analysis without it (like in a traditional applications), however it offers much greater flexibility and extensibility. That is because services used for analysis of the context can be chosen in the real time (when the application is running) from the currently available in the IS services. The other main advantage of this mechanism is that the process of context analysis becomes context-aware itself. Depending on the current situation in the IS it can pick different services and accommodate to current IS capabilities.

Future work will be focused on introducing QoS attributes to the service model and process of service selection. Further, the research will concentrate on creating scenarios of execution of supporting services in an automatic way. It may happen that the IS does not have a service that can directly analyze context data but such analysis might be possible with the usage of some other available services.

The proposed mechanism is not restricted to context analysis. It can be also used to enhance action execution (as they are also based on services). Its usage allows to express actions using semantic description prepared by the user of the CAA which could be then translated into scenario of services execution. Those services could be chosen based on semantic description as well as on interface compatibility (regarding data passed by the user or context data).

References

1. Chaari S, Badr Y, Biennier F (2008) Enhancing web service selection by QoS-based ontology and WS-policy. In: Proceedings of the 2008 ACM symposium on applied, computing, pp 2426–2431
2. Christensen K, Olesen TH, Thomsen LL (2006) Matching semantically described web services using ontologies. Inf Technol Control 35(3A):267–275
3. Hao Y, Zhang Y (2007) Web services discovery based on schema matching. In: Proceedings of the thirtieth Australasian conference on computer science, vol 62, pp 107–113
4. Hsu HJ, Wu SY, Wang FJ (2010) A methodology to developing situation-aware pervasive applications with service oriented architecture, proceedings of the 6th world congress on services, pp 657–660
5. Krawczyk H, Nasiadka S (2011) A new model for context-aware applications analysis and design. In: The fifth international conference on mobile ubiquitous computing, systems, services and technologies, pp 211–217
6. Lee JH, Hashimoto H (2000) Intelligent space, proceedings of the international conference on intelligent robots and systems, pp 1358–1363
7. Lv Q, Zhou J, Cao Q (2009) Service matching mechanisms in pervasive computing environments. In: International workshop on intelligent systems and applications, pp 1–4
8. Maximilien EM, Singh MP (2004) A framework and ontology for dynamic Web services selection. Internet Comput 8(5):84–93
9. Medjahed B, Bouguettaya A (2005) A dynamic foundational architecture for semantic web services. Distrib Parallel Databases 17(2):179–206
10. Paolucci M, Kawamura T, Payne TR, Sycara KP (2002) Semantic matching of web services capabilities. In: Proceedings of the first international conference on semantic web, pp 333–347
11. Yang Z, Chen J, Wu B (2010) A new ontology-based service matching algorithm. In: 6th world congress on services, pp 170–171

Compression of ECG Signals Using a Novel Discrete Wavelet Transform Algorithm for Dynamic Arrythmia Database

Sangeeta Gupta and Sujoy Bhattacharya

Abstract ECG signals play an important role in the primary diagnosis, prognosis and survival analysis of heart diseases. In this paper, a new approach based on the threshold value of ECG signal determination is proposed using Wavelet Transform coefficients. The electrocardiogram signal contains an important amount of information that can be exploited in different manners [1]. Different ECG signals are used to verify the proposed method. A wavelet-based electrocardiogram (ECG) data compression algorithm for dynamic Arrythmia database is presented. The ECG signal is first processed then discrete wavelet is sent to the preprocessed signal. Compression is achieved by variable length based on run length encoding to compress significance and direct binary representation for representing significant coefficients. The proposed algorithm is compared with direct wavelet based compression algorithm represents superior performance.

Keywords Arrythmia · Orientation tree · Denoising factor · Biorthogonal wavelets · Holter records

1 Introduction

Electrocardiograms (ECG) are signals that originate from the action of the human heart. The ECG is the graphical representation of the potential difference between two points on the body surface, versus time. ECG uses information from generation and propagation of electric signals in to the heart to generate clinically relevant

S. Gupta (✉) · S. Bhattacharya
BVRIT, Narsapur, Medak Distt, AP, India
e-mail: sangeeta.gupta@bvrit.ac.in

S. Bhattacharya
e-mail: sujoy.b@bvrit.ac.in

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_78,

information parameters such as time intervals between waves, duration of each wave or composite wave forms and peak amplitudes.

1.1 Arrhythmia

Arrhythmia (ah-RITH-me-ah) is a problem with the rate or rhythm of the heartbeat. During arrhythmia, the heart can beat too fast, too slow, or with an irregular rhythm. Most arrhythmias are harmless, but some can be serious or even fatal. During an arrhythmic attack, the heart may not be able to pump enough blood to the body. Lack of blood flow can damage the brain, heart, and other organs.

In this paper, compression results of ECG signals are shown by using a time-frequency transformation like Discrete Wavelet Transform. Samples of signals are transformed to appropriate groups of transformation coefficients. Almost all coefficients below the determined threshold are rounded to zero values and by inverse transform the similar signal to original one is created. By using run-length coder, consecutive zero value coefficients can be replaced by single value that shows how many consecutive coefficients with zero value exists. In this way small number of coefficients is stored, and compression is obtained. Depending on transform used, different number of coefficients is rounded to zero in different positions, hence the reconstructed signal turns out to be more or less similar to the original one.

1.2 MIT-BIH Arrhythmia Database Directory

The source of the ECG included in the MIT-BIH Arrhythmia Database is a set of over 4000 long-term Holter recordings that were obtained by the Beth Israel Hospital Arrhythmia Laboratory between 1975 and 1979. Approximately 60 % of these recordings were obtained from in-patients.

2 Objectives and Scope

The development of the electrocardiogram was the culmination of scientific effort aimed at perfecting a device conceived for the elucidation of a physiological phenomenon. The morphology of ECG signal has been used for recognizing such variability of heart activity, so it is very important to get the parameters of ECG signal clear without noise. In order to support clinical decision making, reasoning tool to the ECG signal must be clearly represented and filtered, to remove all noises and artifacts from the signal [2]. ECG signal is one of the bio signals that are considered to be non-stationary and this signal needs a lot of hard work for denoising.

An efficient technique for such non-stationary signal processing is the wavelet transforms. The wavelet transform can be used as a decomposition of a signal in the time frequency scale plane. Several methods to enhance ECG signal have been developed. The most widely used is the least mean square adaptive algorithm (LMS). But this algorithm is not able to track the rapidly varying non-stationary signals such as the ECG signal within each heartbeat, which causes excessive low pass bias of mean parameters such QRS complex. We also use bi-orthogonal wavelet transform for ECG signal compression and main ECG parameters estimation.

2.1 ECG Recordings

Based on the conventional criteria on amplitudes and duration of the P-wave, statements are derived regarding left atrial overload, right atrial overload, and prolonged atria conduction (Fig. 1).

Each ECG resting record consists of a header of 1024 bytes and 10 s of ECG raw data. The header is used for acquisition and storage of demographic data including data on the acquisition device which might be needed for construction of a Standard Communications Protocol File (SCP-File). The raw data should be digitized with 500 samples per second and an amplitude quantization of 1.0...5 μV per LSB.

2.2 QRS-T-Evaluation

An adaptive threshold is used (related to the maximum and mean values of the signal), to find the points over this value. After that, the R peaks are selected and stored in a parameter data vector. R-R intervals are known as R-R distances. Q, S

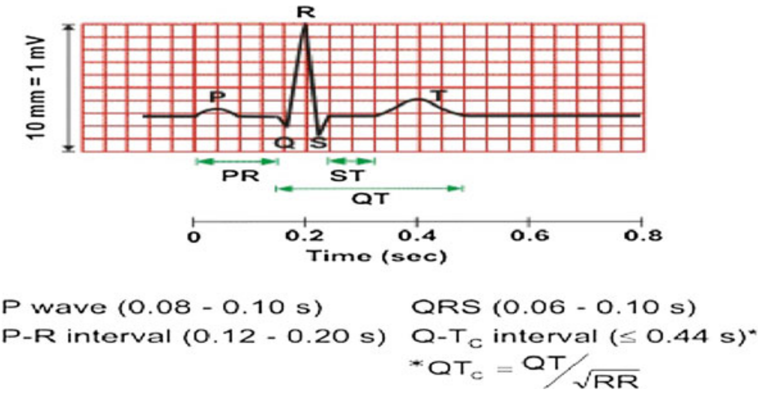


Fig. 1 PQRS time interval

Table 1 PRT detection

Action taken	S (%)	P (%)
R peak detection	>99.2	>99.6
T wave detection	>99.4	>99.5
P wave detection	>98.6	>98.7

points are local minimum points before and after an R wave. The area of the QRS complex can be calculated from the Q-S duration and the value of the R peak. Our test signals contained rhythms from a normal, healthy heart and signals with abnormalities in order to find the main parameters. The used databases were: MIT-BIH Atrial fibrillation database (ADDB), MIT-BIH Supraventricular Arrhythmia Database (SVDB), MIT-BIH Arrhythmia Database (MITDB) and MIT-BIH Normal Sinus Rhythm Database (NSRDB). Over 100 signals were used. The evaluation of this procedure was carried out using the following indicators:

$$\text{Sensitivity (S)} = \text{TP}/(\text{FN} + \text{TP} + \text{FP})$$
$$\text{Positive Predictability (P)} = \text{TP}/(\text{TP} + \text{FP})$$

where TP is the number of true positive detection, FN stands for the number of false negative misdetections and FP is the number of false positive misdetections. The results obtained were compared manually in Table 1 (around 60 beats/signal in the self-created database) and the number of misdetections was noted and the indicators calculated.

3 Discrete Wavelet Transformation

A wavelet is simply a small wave which has energy concentrated in time to give a tool for the analysis of transient, non stationary or time-varying phenomena such as a wave shown in Fig. 2.

A signal as the function of f(t) shown in Fig. 2 can often be better analyzed and expressed as a linear decomposition of the sums & products of the coefficient and function. The set of coefficients are called the Discrete Wavelet Transform (DWT)

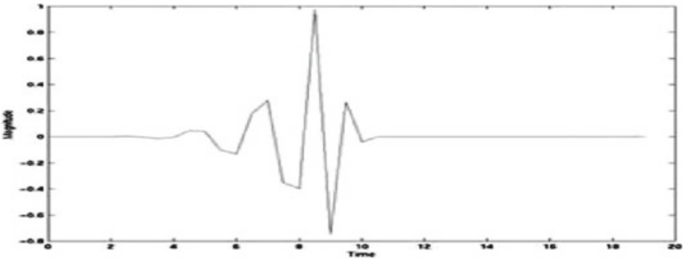


Fig. 2 Wavelet function

of f(t). In the wavelet transform, the original signal (1-D, 2-D, 3-D) is transformed using predefined wavelets [3]. The wavelets can be orthogonal, ortho-normal, or bi orthogonal, scalar or multi wavelets.

3.1 1-D DWT of ECG

The most popular scheme to implement DWT is the convolution-based FIR structure introduced by Mallat in 1988. For a one-dimensional signal, each convolution-based 1D-DWT can be realized using a pair of filters. In the decomposition stage, the original signal x of length L is passed through a low pass filter (LPF) l and a high pass filter (HPF) h, which are called the analysis (decomposition) filter pair [l, h]. Then, the two filtered outputs are down-sampled by 2 to obtain two sub bands, L_1[x] and H_1[x], respectively. In the reconstruction phase, L_1[x] and H_1[x] are up-sampled by 2 and passed through another set of filters called the synthesis (reconstruction) filter pair [$l\sim$, $h\sim$], to reconstruct the signal x.

4 Analyses

We carried out a retrospective analysis of ECG signals recorded during out-of-hospital treatment of adult patients. Four parameters—centroid frequency (FC), peak power frequency (FP), average segment amplitude (SA) and average wave amplitude (WA)—were extracted from the recorded ECG signal immediately before each counter shock and compared with counter shock outcome.

4.1 Clustering

An additional first stage of pre-clustering has been used for removing redundant information. Specific heartbeats have been considered redundant if it has a dissimilarity measure to others.

4.2 Orientation Tree

A tree structure, called “temporal orientation tree”, defines the temporal relationship in the wavelet domain. Every point in layer i corresponds to 2 points in the next layer (i+1), mutually sharing a parent-offspring relationship. This definition is analogous to that of spatial orientation trees in each node that either have no offspring or 2 offsprings. In a typical 1-D signal, most of the energy is concentrated in low frequency bands, so that the coefficients are expected to be better magnitude-ordered as we move downward following the temporal orientation tree to the leaves (terminal nodes).

In digital signal processing, the fast forward and inverse wavelet transforms are implemented as tree-structured, perfect-reconstruction filter banks [4]. The input signal is divided into contiguous, non overlapping blocks of samples called frames and is transformed frame by frame for the forward transform. Within each frame, the input signal is filtered by the analysis filter pair to generate low pass and high pass signals, which is then down sampled by a factor of two. Then this analysis filter pair is applied to the down sampled low pass signal recursively to generate layered wavelet coefficients shown in the resolution.

The number of layers determined the coarsest frequency resolution of the transform and should be at least four for adequate compression [5].

5 Results and Discussions

The frame size of the ECG signal, M, has to be at least 2^{n+1} samples long, where n is the number of levels in the Wavelet Transform.

For each patient, 12 leads (channels) with 6 s of ECG signal per lead were recorded in a database. Every signal is then decomposed using parameters mentioned and threshold analysis is performed. For selected parameters, after reconstruction of the signal in each Lead, compression ratio, maximal absolute difference and PRD is calculated, as it is shown in the Table 2 for DWT with decomposition on 4th level using Daubechies 2 (db2) wavelet function.

In the last row of the Table 2 are shown maximal values of PRD and Max D for that patient, because that is the value of compression which will be obtained if compressed coefficients of all leads are stored in the same file. These values will be called statistical values of compression for selected parameters [6].

Table 2 PRD maximal absolute difference (Max D) and number of zeroes (NZ) for a patient with AIM

ECG	Lead PRD (%)	Max D	NZ (%)
D1	1.55	5	91.61
D2	1.80	4	91.60
D3	1.74	6	91.60
aVR	3.64	3	91.62
aVL	2.25	5	91.60
aVF	1.65	4	91.60
V1	2.16	3	91.60
V2	1.31	6	91.61
V3	1.81	10	91.60
V4	3.37	8	91.61
V5	1.40	10	91.60
V6	2.78	6	91.60
Statistical values	3.6	4	91.6

Table 3 Compression results of ECG signal obtained by using DWT and wavelet packet (WP) by wavelet function and by level of decomposition

Method	Level	db2	db4
DWT	5	91.9	91.71
WP	5	90.18	91.73
WP	6	90.51	92.28
WP	7	90.68	92.44

Reconstruction results are entered in a new table with four additional columns: heart disease id, patient id wavelet function and level of decomposition. For each combination of patient, wavelet function and level of decomposition, a new row is created. There were 10 (patients) 3 (levels of decomposition) 6 (wavelet functions) = 180 rows in a table which represents compression results for 12 (leads) 180 = 2160 ECG.

Two tables of described format are created: one for DWT and another for WP. From these huge tables, regardless of patient, the average compression ratios per combination of wavelet function and level of decomposition are calculated and shown in the Table 3 and appropriate maximal PRD in the Table 4. It is very important to note that the original and reconstructed signals for each lead (ECG signal) had to be visually checked for distortion.

6 Conclusions and Enhancements

The presented method shows a new experimental threshold of wavelet transform coefficients. This threshold value is accomplished experimentally after using a loop of calculating a minimum error between the denoised wavelet sub signals and the original free of noise sub signals. This algorithm performs quite well in terms of both numerical and visual distortion. From the result it is confirmed that the bi orthogonal wavelets tested performed better than the orthogonal wavelets. The DWT wavelet coder appears to be superior to the other wavelet based ECG coders that have appeared in the literature. It has the advantage that no domain specific knowledge or techniques were used in this algorithm. Generally, by using WP, better compression is achieved compared to the DWT.

Table 4 Mean compression results and mean PRD of ECG signals obtained by using M channel cosine-modulated filter bank

Method	Level	db2	db4	db10	Bi- or 4.4
DWT	5	9.9	9.71	9.69	9.7
WP	5	9.27	8.76	8.65	8.79
WP	6	9.38	8.91	8.65	8.95
WP	7	9.39	8.94	8.65	9.02

References

1. Jalaleddine SMS, Hutchens CG, Strattan RD, Coberly WA (1990) ECG data compression techniques—a unified approach. IEEE Trans Biomed Eng 37(4):329–343
2. Bui TD, Chen G (1999) Translation-invariant denoising using Multi wavelets. IEEE Trans Signal Process 46:3414–3420
3. Hilton ML (1997) Wavelet and wavelet packet compression of electrocardiograms. IEEE Trans Biomed Eng 44:394–402
4. Said A, Pearlman WA (1996) A new, fast and efficient image coder based of set partitioning in hierarchical trees. IEEE Trans Circuits Syst Video Technol 6:243–250
5. Strela V, Heller PN, Strang G, Topiwala P, Heil C (1996) The application of multi wavelet filter banks to image processing. IEEE Trans Image Proc, pp. 548–563 (also Technical Report, MIT, Jan. 1996)
6. Soltanian-Zadeh H, Jafari-khouzani K (2002) Multiwavelet grading of prostate pathological images. In: Proceedings of SPIE medical imaging conference, San Diago, CA

Performance Analysis of ETX and ETT Routing Metrics Over AODV Routing Protocol in WMNs

Satish Hatti and M. B. Kamakshi

Abstract Implementation and performance analysis of Expected Transmission Count (ETX) and Expected Transmission Time (ETT) as the routing metric over Ad-Hoc on Demand Distance Vector (AODV) routing protocol are described. Simulation results using NS-2 simulator demonstrate the improved performance of ETX and ETT metric compared to that of Hop-Count metric in static node topology. Minimizing the Hop-Count maximizes the physical length of each link in a path, which is likely to minimize signal strength and maximize the loss ratio. One contribution of this paper is to quantify these effects. The performance is evaluated in terms of throughput, packet loss rate, and end to end delay for the above three metrics. For long paths the throughput improvement is doubled for ETX and ETT metric when compared to Hop-Count metrics, suggesting that ETX and ETT metrics will become more useful as network grows larger and paths become longer.

1 Introduction

Today, there is a consensus that routing metrics that do not consider physical variations are not suitable for Wireless Mesh Networks (WMNs). Metrics unaware of link quality cannot guarantee reasonable stability and acceptable loss rates. Therefore, routing in wireless mesh networks has evolved by designing algorithms that take wireless medium conditions into account. Thus, the recently proposed metrics reflect various physical-layer characteristics, such as loss probability and transmission rate.

In ad hoc networks, the most used metric is Hop-Count, which is convenient for ad hoc because frequent link breakages result from the mobility of users.

S. Hatti (✉) · M. B. Kamakshi
Department of Telecommunication Engineering, R.V.C.E, Bangalore 560059, India
e-mail: satish034@gmail.com

M. B. Kamakshi (✉)
e-mail: kamakshimb@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_79,

Much of the recent work in ad-hoc routing protocols for wireless networks [1] have focused on coping with mobile nodes, rapidly changing topologies and scalability. On the other hand, less attention has been paid in finding high-quality paths in the lossy wireless links as WMN routers are usually stationary, routing metrics that reflect link quality variations are feasible.

The Hop-Count metric protocols typically use only links that deliver routing probe packets. This approach implicitly assumes that links either work well or don't work at all.

The Hop-Count metric chooses the path arbitrarily among the different paths of the same minimum length, regardless of the often large differences in throughput among those paths, and ignoring the possibility that a longer path might offer higher throughput. One approach to fix this problem is to mask transmission errors in two ways,

- Resend the lost packets, using the 802.11 ACK mechanisms. However, retransmission does not make lossy links desirable for use in paths: the retransmissions reduce path throughput and interfere with other traffic, or
- Increase Hop-Count routing with a threshold that ignores lossy links, but a lossy link may be the only way to reach a certain node, and there might be significant loss ratio differences even among the above threshold links.

The solution proposed and evaluated in this paper is the ETX and ETT metric. The ETX metric incorporates the effects of link loss ratios, asymmetry in the loss ratios between the two directions of each link, and interference among the successive links of a path. ETT metric reproduces link quality conditions of ETX metric and also considers the packet size and physical transmission rates in establishing the link from source to the destination.

2 Routing Metrics

This section describes the design of the metric. The metric's overall goal is to choose routes with high end-to-end throughput.

2.1 Working of ETX Metric

The primary goal of the ETX design is to find paths with high throughput, despite losses. The ETX of a link is the predicted number of data transmissions required to send a packet over that link, including retransmissions. The ETX of a route is the sum of the ETX for each link in the route.

The ETX of a link is calculated using the forward and reverse delivery ratios of the link. The forward delivery ratio D_f is the measured probability that a data packet successfully arrives at the recipient; the reverse delivery ratio D_r is the probability that the ACK packet is successfully received. The expected probability that a transmission is successfully received and acknowledged is $\mathrm{D_f} \times \mathrm{D_r}$. The expected number of transmissions is:

$$ETX = \frac{1}{D_f \times D_r} \tag{1}$$

The delivery ratios $\mathrm{D_f}$ and $\mathrm{D_r}$ are measured using dedicated link probe packets. Each node broadcasts link probes of a fixed size, for an interval of a window size w (10 s in this implementation). Because the probes are broadcast, 802.11 do not acknowledge or retransmit them. Every node remembers the probes it receives during the last w, allowing it to calculate the delivery ratio from the sender at any time t as:

$$r(t) = \frac{count(t - w, t)}{w/\tau} \tag{2}$$

where, $count(t - w), t$: number of probes received during the interval w, w/τ: expected number of probes that should have been received.

2.1.1 ETX Metric Implementation

Routers periodically broadcast small size probes to estimate $\mathrm{D_f}$ and $\mathrm{D_r}$. The ETX computation considers both forward and reverse directions because of data and ACK frame transmission. The ETX of a route is the sum of the link metrics. AODV accumulate the route metric as it forward updates. The chosen route is the one with the lowest sum of ETX values along the route to the destination.

2.1.2 The Shortcomings of ETX

The shortcomings of ETX metric are listed as follows

- ETX assumes that radios have a fixed transmit speed, and that congestion is not a factor.
- ETX also assumes that delivery rates are the same regardless of packet size, which is probably inaccurate.
- IEEE 802.11 broadcast frames are sent at the network basic physical rate and probes are usually smaller than data packets.

To cope with these problems, the ETT [2, 3] metric was proposed. The ETT metric estimates the time a data packet needs to be successfully transmitted on a link.

Table 1 List of system parameters considered in simulation

Parameters	Value
Channel type	Wireless channel
Propagation model	Two ray ground
MAC layer	IEEE 802.11
Routing protocol	AODV
Traffic model	UDP/CBR
Total simulation time	100 s and 500 s
Number of nodes	25

2.2 ETT Metric Implementation

In AODV, each node periodically broadcasts HELLO messages in order to know the state of local links. Each node lists its neighbours in HELLO messages and consequently, a node is aware of its two-hop neighbours (Table 1).

The ETT metric computation is performed using the method proposed by Draves et al. [4]; the ETT metric can be calculated by adjusting the ETX metric according to the packet size and the transmission capacity of the link. It is calculated using the expression as,

$$ETT = ETX \times S \big/ B \tag{3}$$

where S is the packet size and B is the link capacity.

To estimate B, nodes use the packet-pair technique [4]. In this technique, two back-to-back probes, one small followed by a large one, are sent to each neighbour. Each neighbour then measures the inter-arrival time between the two packets and reports it back to the sender of the probes. Upon receiving a predefined number of delay samples, the sender estimates the capacity of the link by dividing the size of the larger probe by the smallest delay sample obtained. Similarly to ETX, the chosen route is the one with the lowest sum of ETT values.

3 Results and Analysis

The performance of ETT, ETX, and Hop-Count metrics over AODV routing protocol is compared considering the three different scenarios, the results and the performance analysis is discussed in each cases.

3.1 Simulation and Performance Analysis of Scenario A

I. **Simulation set up for Scenario A**
Scenario A is a WMN with four lossy nodes (i.e. 6, 8, 16 and 18) which all have 50 % incoming loss ratio as shown in Fig. 1 There is only one communicating

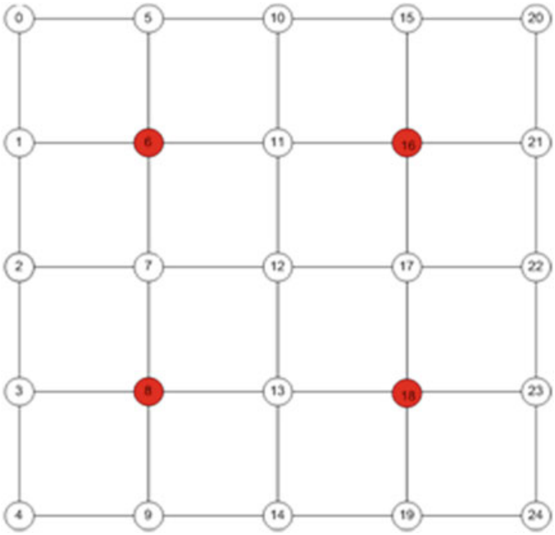


Fig. 1 Illustration of four lossy nodes in Scenario A (nodes 6, 8, 16 and 18 are lossy nodes)

node pair in Scenario **A**, which is from nodes 0 to 24. Scenario **A** is similar to the scenarios used in previous works [4–6].

II. **Simulation result for Scenario A**
See Fig. 2.

III. **Performance analysis of Scenario A**
In Scenario A, the performance of the ETX metric will be better compared to the Hop-Count metric. The lower throughput in Hop-Count indicates that it will also use the lossy nodes in transmission. The higher throughput in ETX proves that it will choose the path with higher packet delivery ratio avoiding the lossy nodes.

3.2 Simulation and Performance Analysis for Scenario B

I. **Simulation set up for Scenario B**
Scenario B is used to simulate a WMN with multiple server-to-client traffic. Consequently, in this scenario there are one source and twenty-four sinks. Node 0 as a server transmits packets to the rest of nodes simultaneously, with CBR traffic over UDP transport. The number of lossy nodes is uniformly distributed between 0 and 12.

II. **Simulation Results for Scenario B**
See Fig. 3.

III. **Performance analysis for Scenario B**

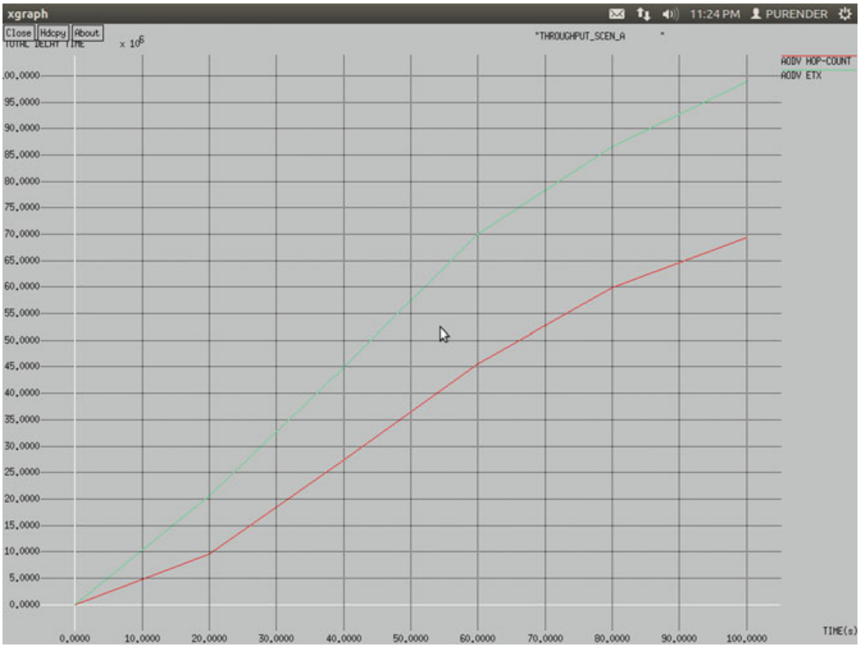


Fig. 2 Throughput ‘for ETX versus Hop-Count metrics in Scenario A

In Scenario B, the performance of the ETX metric will be better compared to the Hop-Count metric. It remains same throughout the execution interval of 100 s.

3.3 *Simulation and Performance Analysis for Scenario C*

3.3.1 Simulation Set Up for Scenario C

5*5 grid topology with 25 nodes as shown in the Fig. 1 is considered for performance analysis. All the 25 nodes are considered to be lossless.

3.3.2 Simulation Results for Scenario C

I. **Simulation Results for throughput in Scenario C**
See Fig. 4.

II. **Simulation Results for Packet loss ratio in Scenario C**
See Fig. 5.

III. **Simulation Results for End-to-End Delay in Scenario C**

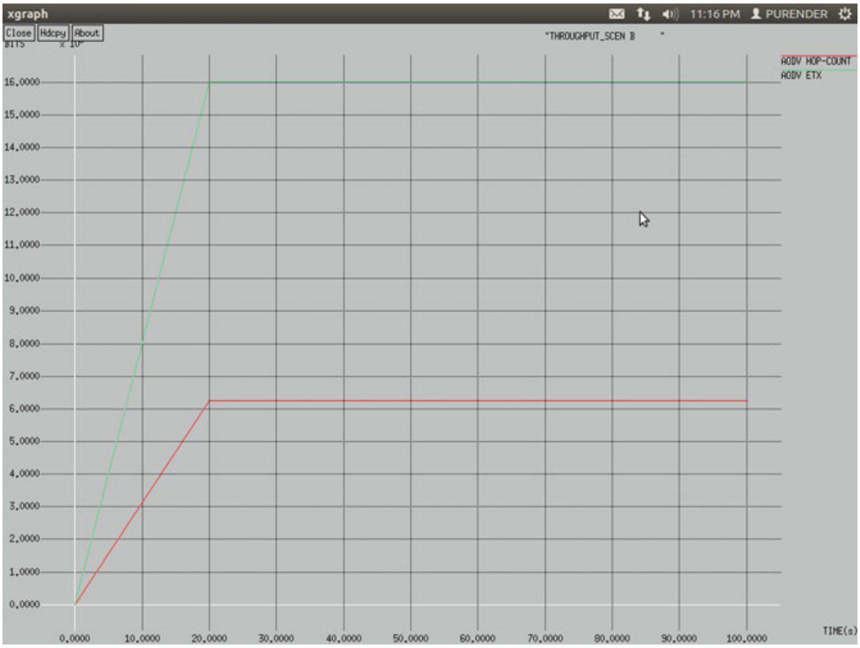


Fig. 3 Throughput for ETX versus Hop-Count metrics in Scenario B

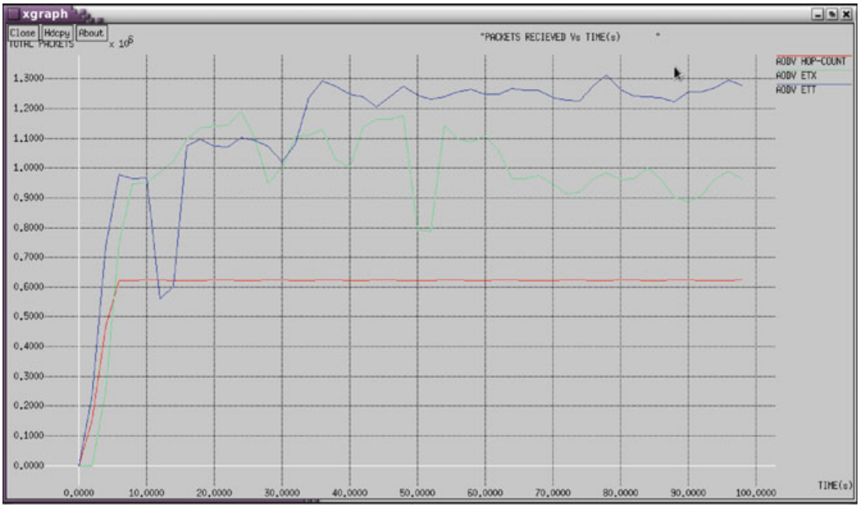


Fig. 4 Throughput for ETT, ETX and Hop-Count metrics in Scenario C

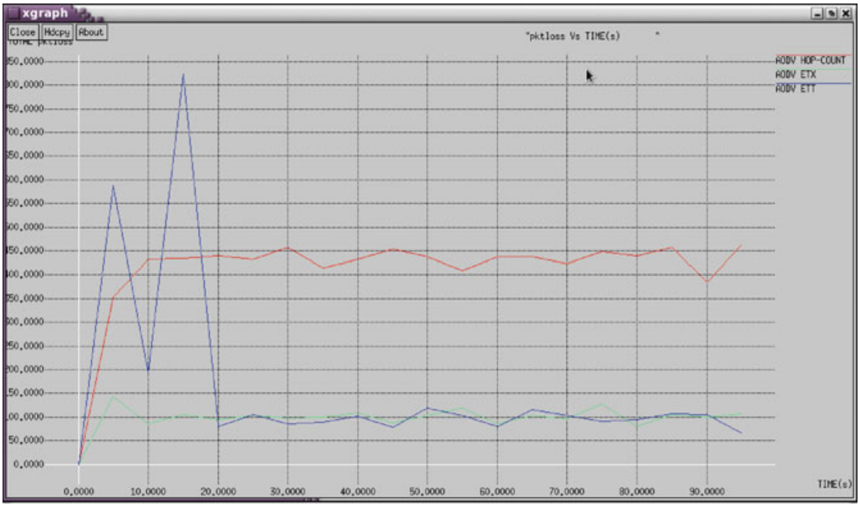


Fig. 5 Packet loss ratio for ETT, ETX and Hop-Count metrics in Scenario C

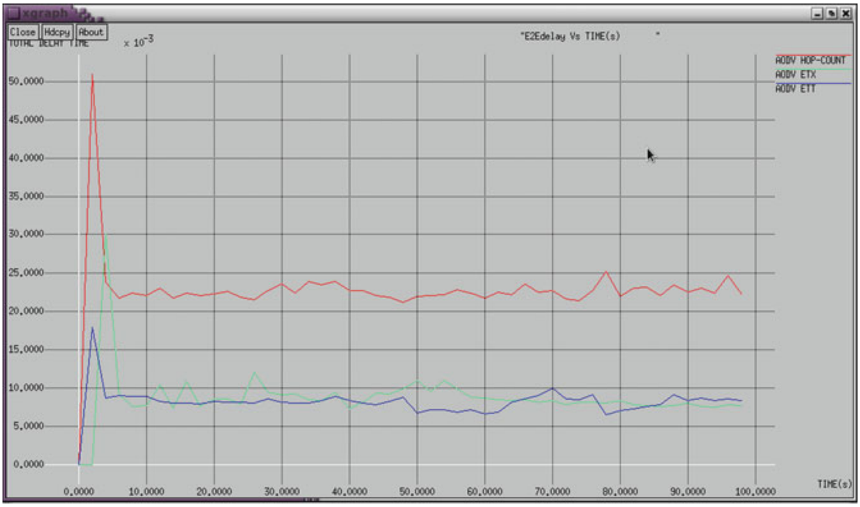


Fig. 6 End-to-End delay for ETT, ETX and Hop-Count metrics in Scenario C

Average end to end delay is defined as the average time between the moment a data packet is sent by a source and the moment the sink receives the data packet. This metric defines the freshness of data (Fig. 6).

3.3.3 Performance Analysis of Scenario C

I. **Throughput**
ETX metric performs better than Hop-Count metrics. This is because the required number of retransmission of packets will be lesser in case of ETX as it considers the path with higher delivery ratio. ETT metric will perform better to ETX metric.

II. **Packet loss ratio**
ETX metric performs better than Hop-Count metrics. This is because the ETX as it consider the path with higher delivery ratio. ETT metric will perform better to ETX.

III. **End-to-End Delay**
ETX and ETT metrics performs better than Hop-Count metrics. This is because the transmission failure is lesser in case of ETX and ETT metrics as it considers the path with higher packet delivery ratio.

4 Conclusions

The experiment shown the ETT metric has the higher Throughput, lower packet loss rate and the lower End to End delay time among the analyzed metrics.

Conclusions based on the results obtained in all scenarios are listed below:

- In case of lossy links the network performance is improved using ETX or ETT metric than Hop-Count.
- ETT metric considers the packet size and Bandwidth in addition to the link quality hence the performance of the ETT will be better compared to the ETX metric.
- The result has shown, for long paths the throughput improvement is doubled for ETX and ETT metrics when compared to hop count, suggesting that ETX and ETT metrics will become more useful as network grows larger and paths become longer.

5 Future Scope

The below metrics can be implemented by considering additional link qualities to overcome non-isotonicity and inability to capture interference.

IBETX metric is interference and bandwidth adjusted *ETX* (*IBETX*) for wireless multi-hop networks.

WCETT metric for routing in multi-radio, multi-hop wireless networks, captures the intra-flow interference of a path since it essentially choose paths that have more diversified channel assignments on their links and hence lower intra-flow interference.

The *MIC* metric, improves *WCETT* by solving its problems of non-isotonicity and the inability to capture inter-flow interference.

References

1. Perkins CE, Royer EM (1999) Ad hoc on-demand distance vector routing. In: Proceedings of Mobile Computing Systems and Applications
2. Broch J, Maltz DA, Johnson DB, Hu Y-C, Jetcheva J (1998) A performance comparison of multi-hop wireless ad hoc network routing. http://www.monarch.cs.cmu.edu/
3. Draves R, Padhye J, Zill B (2004) Routing in multi-radio, multi-hop wireless mesh networks, pp 114–128. doi:10.1145/1023720.1023732
4. Draves R, Padhye J, Zill B (2004) Comparison of routing metrics for static multi-hop wireless networks. ACM SIGCOMM Comput Commun Rev 34(4):133–144. doi:10.1109/WD.2008.4812900
5. De Couto D, Aguayo D, Bicket J, Morris R (2005) A high-throughput path metric for multi-hop wireless routing. Wireless Netw 11(4):134–146. doi:10.1007/s11276-005-1766-z
6. Yun TT (2007) Experimental analysis on route oscillations in ETX within wireless meshes networks. Dissertation, University of Sydney

Optimized CPU Frequency Scaling on Android Devices Based on Foreground Running Application

Tanuj Mittal, Lokesh Singhal and Divyashikha Sethia

Abstract When it comes to portable devices, battery life is a very crucial aspect. Battery life of a mobile device can be improved by several techniques. One such technique is the effective management of the operating voltage and frequency of the processor. Linux provides a utility, named cpufreq, to dynamically scale CPU frequency on the fly. Since Android operating system is based on Linux, the cpufreq utility can be used to scale the CPU frequency of the portable devices. To decide when and how much frequency should be scaled, cpufreq governors are used. This paper proposes the basis for a new cpufreq governor called appspace, to improve battery life by considering not only instantaneous CPU usage, but also foreground running application on the device.

1 Introduction

Over the last decade, more and more portable devices have entered the market. Along with the many benefits these devices bring, there are some major concerns like longer battery life that can enhance end-user experience. To obtain a longer battery life, in addition to technological advances in battery technology, we also need to focus in making the software as efficient as possible. Common techniques used to optimize battery on portable devices include changing display brightness level, efficiently managing wireless services (like Wi-Fi, Bluetooth, GPS, Cell-radio), killing or suspending unwanted background processes etc. [1, 2].

Modern CPU architecture allows the processor to operate at multiple clock frequencies. A majority of the Android devices provide this feature. By switching to a lower operating frequency, the voltage required by the CPU may decrease [3]. If done effectively, this method proves to be very useful in lowering down average power

T. Mittal (✉) · L. Singhal · D. Sethia
Department of Software Engineering, Delhi Technological University, Delhi, India
e-mail: tanuj.183@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_80,

consumption by the processor and hence saving battery power. Android operating system runs on Linux based kernel. The Linux kernel provides a useful utility, named cpufreq, to manage the CPU frequency. It allows configuration of CPU frequency at hardware level according to a cpufreq governor (more about this is discussed under Sect. 2). All the governors have certain parameters which can be tuned for a particular system. These parameters are present in the form of sysfs[1] files.

Among many factors triggering CPU frequency scaling, CPU usage is prominent. It gives an estimate to the overall processing needs of the system at an instant. However, the foreground running application helps to identify the maximum frequency level that may be needed by the system throughout application's runtime. We have investigated this for the first time for improving power optimization by CPU frequency scaling. When considered together, with the help of these factors, we can determine when to scale the frequency and to what extent (discussed under Sect. 3). By setting the maximum limit on the frequency scaling range of the processor, based on foreground running application, unnecessary jumps to unneeded frequency levels can be eliminated. This proves to be useful, because different applications have different processing requirements and some of them can even function effectively at low frequencies.

The remaining part of the paper is broadly divided into six sections. Section 2 discusses dynamic frequency scaling and existing methods to do so. Sections 2.1 and 2.2 provide information about existing frequency scaling models present in Linux kernel. The new model proposed is discussed under Sect. 3. We demonstrate the functioning of this model using a demo application under Sect. 4. Sections 5 and 6 provides results and related work. Section 7 concludes the work and provides information regarding future work that shall be done on the lines of the proposed frequency scaling model.

2 Dynamic CPU Frequency Scaling

Dynamic voltage and frequency scaling (DVFS) is a commonly-used power-management technique where the clock frequency of a processor is decreased to allow a corresponding reduction in the supply voltage [4]. This technique is used in portable devices like laptops, mobiles, tablets etc. where battery life is of high priority. The relation for the dynamic power (switching power) dissipated by the chip is given in Eq. (1)

$$\mathrm{P} = \mathrm{C}\ \mathrm{V}^2\mathrm{f} \tag{1}$$

where P is the dynamic power (switching power) dissipated by the chip, C is the capacitance of the transistor gates (which depends on feature size), f is the operating frequency and V is the supply voltage [5].

[1] Sysfs is a virtual file system provided by Linux 2.6 and above. It exports information about devices from the kernel device model to user space, and is also used for configuration.

The voltage required for stable operation is determined by the frequency at which the circuit is clocked, and can be reduced if the frequency is also reduced. This can yield a significant reduction in power consumption because of the V^2 relationship shown in (1). The CPU usage is calculated in terms of jiffies.[2] We can process data from /stat/proc file to calculate the number of jiffies at any instant, which in turn can be used to estimate current CPU usage [6]. If the number of jiffies at an instant is high, the processing requirement of the system is also high [7].

2.1 Cpufreq

The Android kernel, which is based on Linux, provides an interface to modify CPU frequency dynamically using cpufreq. Cpufreq is a hardware service that probes and configures CPU frequency daemon module of kernel. It works on cpufreq governors present in the kernel itself. These governors direct the cpufreq daemon about when and how much to scale frequency. Frequencies can be scaled automatically depending on the system load, in response to ACPI (Advanced Configuration and Power Interface) events, or manually by userspace programs

2.2 Cpufreq Governors

Governors are pre-configured power schemes for the CPU. They must be loaded as a kernel module in order to be seen by userspace programs. One may load as many governors as desired (only one will be active on a CPU at any given time). Governors such as performance, powersave, userspace, ondemand, conservative and interactive are shipped by default in Linux [8].

Unlike other governors, userspace governor allows the user, or any userspace program running with UID (user ID of the process) root, to set the CPU to a specific frequency by making a sysfs file scaling_setspeed available in the CPU-device directory (/sys/devices/system/cpu/cpu0/cpufreq, in case of kernel 2.6.35). This way, a userspace program can also change the frequency without having to create any new kernel modules. We have used this governor to change frequency from our test application, which is based on the proposed model named appspace (detailed in Sect. 4).

Ondemand governor sets the frequency depending on the current CPU usage. It sets the frequency corresponding to the instantaneous CPU usage. It is best suited for environments, like desktops, where energy consumption is of medium priority and system latency doesn't matter much [9]. By default, most of the Android devices ship with ondemand governor as default. Therefore, we compare our proposed model, appspace, with ondemand in order to calculate the battery savings achieved.

[2] A jiffy is a proverbial short amount of time, which is 1/100 s on most CPUs (depends on highest clock rate of CPU).

3 Proposed Work

We have proposed a new governor which considers CPU usage as well as the foreground running application to help to identify the maximum frequency level that may be needed by the system throughout application's runtime. All the applications are categorized broadly on the basis of their processing needs. This categorization can be done dynamically (on Android using Android Market APIs) whenever an application is launched for the first time and the information can be stored for future use. The categories are mapped to the maximum permissible frequency limit and it is used to set the corresponding frequency during runtime. The governor is developed at a user level in the form of an application called appspace.

Suppose, the available CPU frequencies on a device are F[0], F[1], F[2],...F[n − 1], where F[0] is the minimum available frequency and F[n − 1] is the maximum available frequency on a device. Further, suppose that the new foreground running application belongs to a category with maximum permissible frequency F[y] (where $0 \leq y \leq n - 1$). Also, SAMPLING_RATE is the probe time to calculate CPU usage at a particular instant; UP_THRESHOLD and DOWN_THRESHOLD are the threshold CPU usages to trigger scaling of the frequency up and down respectively. As soon as a new application is launched, the frequency is set to F[y] i.e. the maximum permissible frequency of the category. Now, according to SAMPLING_RATE, CPU usage will be calculated every probe cycle. Assuming, F[t] is the current frequency in a particular probe cycle (where F[0] ≤ F[t] ≤ F[y]) and CPU_USAGE is the CPU usage in the cycle, then one of the following cases may arise:

- CPU_USAGE > UP_THRESHOLD: In this case, CPU frequency will be set to F[y] i.e. the maximum permissible frequency of the category.
- CPU_USAGE < DOWN_THRESHOLD: In this case, CPU frequency will be set to F[t − 1], if $0 \leq t - 1 \leq y$, otherwise it will remain F[t].
- DOWN_THRESHOLD ≤ CPU_USAGE ≤ UP_THRESHOLD: In this case, the frequency will remain F[t].

Whenever an application is first launched, the processing requirement is a little higher than average because of initial resource loading. This could lead to stuttering user interface and system slow-downs. To prevent this, the frequency is set to the maximum of its category at first, and eventually scaled down to lower frequencies as the processing requirements of the application drop. At anytime, when the CPU usage crosses the threshold, the frequency is again set to the maximum to maintain lag free experience. This removes any latency that could occur due to slow processing state of CPU. But as soon as the CPU usage drops, the CPU frequency is immediately scaled down. This technique not only improves the overall responsiveness of the system but also keeps the CPU in a minimum possible frequency state. Here, we are decreasing the frequency linearly whenever CPU usage is below a threshold; we can also use an algorithm to map the CPU usage to the corresponding frequency for faster and more efficient frequency scaling. This will enable the CPU to jump from high frequency levels to low frequency levels directly, if possible, and hence conserve even more battery power. The concept of setting limit on maximum frequency can be extended to different device states. The Android device can be in different states on the basis of screen being locked or the display being off. If this is the case, the device requires low or no processing at all, hence similar to low demanding applications category.

4 Testing

The proposed model, as discussed under Sect. 3, was tested on an Android device by developing a test application. The application works on the same principle as proposed cpufreq governor, appspace, but on a user level.

It acts as an application based governor and tracks the user activity while making decisions to ramp the frequency up or down. It works on an Android distribution with userspace governor available in the kernel. It also requires the user to have root access to modify some sysfs files. The application consists of a background service to continuously track new application being launched, a receiver to act upon certain actions and a database to map applications to corresponding categories. It modifies frequency by using userspace governor, which sets the CPU frequency to the value contained in the sysfs file named scaling_setspeed. The application writes desired frequency to the file whenever needed. Due to lack of availability of resources, LG Optimus One P500 was used as the testing device with Linux kernel 2.6.35 and Android 2.3.7 running [10].

Two frequency scaling models, ondemand and appspace, were tested. Since ondemand is the default governor in many of the Android devices available in the market, the testing was done against it to demonstrate battery savings. Other governors, like powersave and performance, are not meant for maintaining a balance between good user experience and long battery life. Hence their testing was not necessary. Our testing application launched the applications automatically and each application was used for 5 min for both the governors separately. This ensured the consistency of the results. In total, 6 applications were used during testing, Angry Birds (high demanding), Browser, Messaging (medium demanding) and Dropbox, The Hindu Newsfeed, Calendar (low demanding). All these applications were either pre-installed or downloaded from Android Market.

During testing, certain parameters, including battery voltage, were recorded in a file for every 2 s over a period of 30 min for each governor. Since the results for voltage were very erroneous and fluctuating, the drop in battery voltage was calculated by subtracting average voltage from starting voltage. Then the percentage drop in battery voltage was calculated. The results are presented in Table 1. The UP_THRESHOLD has been taken as 90 % and DOWN_THRESHOLD as 50 % for testing .

5 Results

When compared to ondemand, the proposed model, appspace, results in higher battery savings while maintaining smooth user experience. It can be inferred from Fig. 1 that ondemand results in a higher rate of decrease in battery voltage whereas appspace results in a comparatively lower rate of decrease in battery voltage for most of the time. According to Table 1, the battery voltage drop is 52.63 % less when using appspace as compared to when ondemand is used.

When category 1 applications, i.e. high demanding applications are used, there is hardly any difference in frequencies set by ondemand or appspace This is because the CPU usage is close to 100 % most of the time, hence the frequencies set are high. In category 2 and category 3 applications, i.e. medium demanding and low demanding, there is a noticeable difference in the frequencies set by both the governors. Figure 2 shows the percentage of time the CPU operates at various frequencies for ondemand and appspace. It can be inferred from these graphs that CPU operates at lower frequencies for most of the time when using appspace as compared to when ondemand is used.

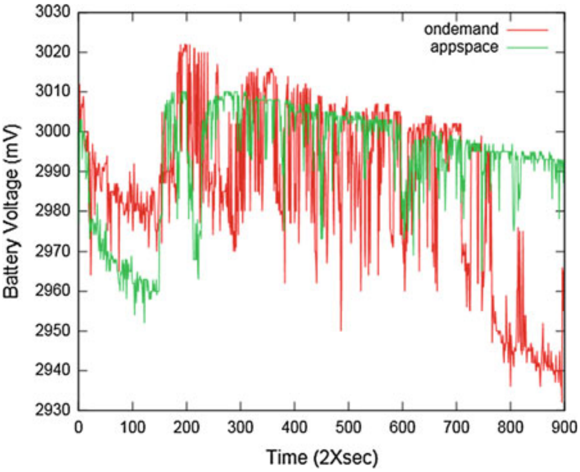


Fig. 1 Battery voltage comparison

Table 1 Testing results

Governor	Starting voltage (mV)	Average voltage (mV)	Drop in voltage (mV)
Ondemand	4001	3986.38	14.62
Appspace	3953	3946.07	6.93

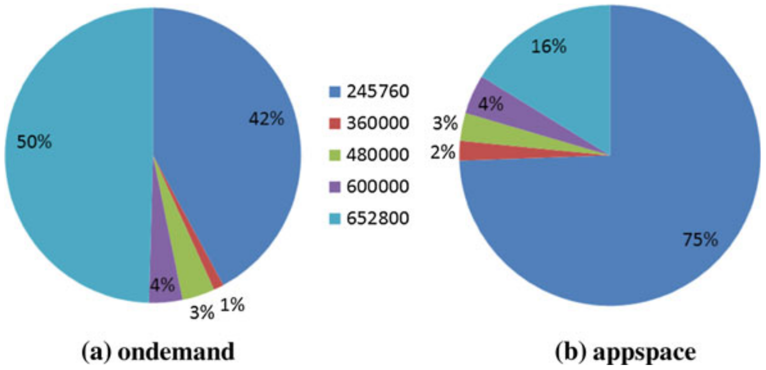


Fig. 2 Time spent by CPU at different frequencies (in kHz). **a** ondemand, **b** appspace

6 Related Work

Traditionally, Linux kernel uses cpufreq governors, which are discussed under Sect. 2.2, for dynamically adjusting CPU frequency. These governors have been adapted over the time for usage on mobile devices. Most of the Android devices ship with ondemand governor as default. Many recent governors, like ondemandx, minmax, smartass, scary, lagfree, savagedzen, lazy, lionheart, intellidemand and hotplug, have emerged on the ideas of existing governors which primarily consider CPU usage [11, 12]. Other attempts to minimize battery usage have been made by predicting the critical speed required by the system to complete a task [13]. But our proposed model gives equal importance to the foreground running application and improves the logic of dynamic scaling even further. By doing this, a desired balance between battery life and performance is achieved which is not possible with other governors.

7 Conclusion & Future Work

The testing reveals that when compared to ondemand governor, which is default in most of the Android devices, the appspace results in higher battery savings without sacrificing the user experience. Since the application itself provides good amount of battery power savings, the kernel level cpufreq governor based on the same model will drastically increase the battery life with almost zero latency. An application to manually configure the categories of applications at user level would complement the governor. Hardware testing along with software testing can provide much more accurate results and help to determine the pros and cons of the algorithm accurately. The scalability of the algorithm can be increased if detection of category of application is possible dynamically (probably by using Android Market APIs). By leveraging the meta-data of currently running applications on a system, the Operating System can greatly enhance the user experience as well as use the resources more efficiently. This can lead to a new class of Operating Systems, which can fine tune various properties of the system, during runtime, by utilizing the meta-data of running applications.

References

1. Welch GF, University of North Carolina, Chapel Hill (1995) A survey of power management techniques in mobile computing operating systems. ACM SIGOPS Oper Rev 29(4):47–56
2. Vallina-Rodriguez N, Hui P, Crowcroft J, Rice A (2010) Exhausting battery statistics, In: Proceedings of the second ACM SIGCOMM workshop on networking, systems, and applications on mobile handhelds, MobiHeld '10, New York
3. Mallik A, Lin B, Memik G, Dinda PA, Dick RP (2006) User-driven frequency scaling. IEEE Comput Archit Lett 5(2):16
4. Sueur EL, Heiser G (2010) Dynamic voltage and frequency scaling: the laws of diminishing returns. In: Proceedings of the 2010 international conference HotPower'10 on power aware computing and systems
5. Pouwelse J, Langendoen K, Sips H (2001) Dynamic voltage scaling on a low-power microprocessor. In: Proceedings of the 7th annual international conference on mobile computing and networking, ACM SIGMOBILE 2001
6. Linux manual page (2007), proc. http://linux.die.net/man/5/proc
7. Choi K, Soma R, Pedram M, University of Southern California, Los Angeles (2004) Dynamic voltage and frequency scaling based on workload decomposition. In: Proceedings of the 2004 international symposium ISLPED '04 on Low power electronics and design
8. Linux kernel documentation (2008), [Linux/kernel/git/torvalds/Linux-2.6.git]/Documentation/cpu-freq/governors.txt
9. Pallipadi V, Starikovskiy A (2006) The ondemand governor. In: Proceedings of the Linux, symposium, vol 2
10. LG optimus one P500 specifications (2010), Gsmarena. http://www.gsmarena.com/lg_optimus_one_p500-3516.php
11. Erasmux, sources of minmax, smartass, smartass2, interactivex governors, Github. https://github.com/erasmux/hero-2.6.29-flykernel/tree/master/drivers/cpufreq
12. Imoseyon (2011) CPU Governors. http://www.imoseyon.com/2011/10/cpu-governors.html
13. Liang Y, Lai P, Chiou C (2010) An energy conservation DVFS algorithm for the android operating system. J Convergence 1:93–100

Message Efficient Ring Leader Election in Distributed Systems

P. Beaulah Soundarabai, J. Thriveni, H. C. Manjunatha, K. R. Venugopal and L. M. Patnaik

Abstract Leader Election Algorithm, not only in distributed systems but in any communication network, is an essential matter for discussion. Tremendous amount of work are happening in the research community on election as network protocols are in need of co-ordinator process for the smooth running of the system. These so called Coordinator processes are responsible for the synchronization of the system otherwise, the system loses its reliability. Furthermore, if the leader process crashes, the new leader process should take the charge as early as possible. New leader is one among the currently running processes with the highest process id. In this paper we have presented a modified version of ring algorithm. Our work involves substantial modifications of the existing ring election algorithm and the comparison of message complexity with the original algorithm. Simulation results show that our algorithm minimizes the number of messages even in worst case scenario

Keywords Election · Coordinator · Message complexity · Ring algorithm · Distributed system

P. B. Soundarabai (✉)
Department of Computer Science, Christ University,
Hosur Main Road, Bangalore, India
e-mail: beaulah.s@christuniversity.in

J. Thriveni, H. C. Manjunatha and K. R. Venugopal
Department of Computer Science and Engineering,
University Visvesvaraya College of Engineering Bangalore University,
Bangalore, India

L. M. Patnaik
Indian Institute of Science, Bangalore, India

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_81,

1 Introduction

Leader Election is a vital and fundamental problem in distributed systems and in any communication network. Distributed Systems is a collection of heterogeneous systems which interact with each other through messages. The main objective of Distributed System is, though there are heterogeneous systems in the network, it creates a single system image or uniprocessor image to the user, through various transparency metrics. The communication between the processes is achieved by exchanging messages. The software of the Distributed System is tightly coupled and the processes of the system coordinate with each other. They have lots of resources in common and so mutual exclusion algorithms are used to take care of the critical regions. While they wait for the common resources, they might end up in a deadlock. Deadlock detection and prevention algorithms should keep an eye on the resources and if there is a deadlock, wound-wait or wait-die algorithms are used to kill the eldest or the youngest process to remove the deadlock.

The replicated data management, group communication, atomic commit protocols, etc needs the process coordination. All the above stated protocols need a particular process among the group, to be the leader and have the control over the situation. In general, the process with the highest process id is the coordinator or the leader. Any process, which satisfies the rule can become the leader and at any point of only one process should be the leader.

Motivation: In Wireless Networks like Wireless LAN, Satellite Oriented Services Cellular Phones, the mobile systems are subject to loss of messages or the data and the mobile host can crash or can be down for some time. Electing a leader process is a basic operation which happens in the system very often. In mobile ad hoc network, we always face node failure or the process crash. Even during the time of leader process crash, there should be a new leader process available, without wasting of time and the number of messages which are exchanged.

Contribution: In this paper, we have proposed a modified ring algorithm with minimized number of messages. The main objective of the algorithm is the fault tolerance. The proposed algorithm minimizes the number of messages in election process thereby reducing the time taken to elect the new leader as compared with the existing election algorithms.

Organization: The remainder of this paper is organized as follows: Sect. 2 reviews the related work; Background of Ring Election Algorithms is available in Sect. 3. The Efficient Ring Election Algorithm is proposed in Sect. 4; Sect. 5 deals with the Performance Evaluation; Conclusions are presented in Sect. 6.

2 Related Work

Sung Hoon Park [1] has proposed a concept called Failure Detector which works as an independent module with a function that detects crash and recovery of a node in a system. This report can be given to any process at request. The author modified the bully algorithm using the failure detector. The performance of the system goes down because of the overhead of Failure Detector. And as the Failure Detector is the centralized component, it has the problem of single source failure and also creates bottleneck situation to access the module.

Sandipan Basu [2] has addressed the issue of bully algorithm and proposed his modified algorithm. In the original bully algorithm, when the leader process is crashed, immediately the new leader is elected. But, if the old leader process comes back, it once again initiates the election. The author suggests that there need not be another election, instead, the old leader process can accept the new leader process by sending the new request of *who the leader is*, to its neighbor. In the next round of election, it can try becoming the leader.

Muhammad Mahbubur Rahman et al. [3] have proposed a modified bully election algorithm. In their paper, they say that the bully algorithm has O(n^2) messages which increases the network traffic. In the worst case, there will be *n* number of elections can occur in the system which again in tern will yield in a heavy network traffic. They have proposed the same algorithm but with Failure Detector, Helper processes to have unique election with the Election Commission.

Chang Young Kim et al. [4] have proposed the election protocol for reconfigurable distributed systems which again was based on bully election algorithm. The actual election is run by the base stations making the protocol, energy efficient. The protocol is independent from the overall number of mobile hosts and the data structures required by the algorithm are managed at the base station, making the protocol scalable.

M. S. Kordafshari et al. [5] discussed the drawback of synchronous bully algorithm and modified it with an optimal message algorithm. The authors have tried to reduce the number of elections happening in the classical bully algorithm. The proposed algorithm has only one election at any point of time, which brings down the number of messages being exchanged drastically. Sepehri M. et al. [6] have dealt with the distributed leader election algorithm for a set of processes connected by a tree network. The authors have proposed a linear time algorithm using heap structure using reheap up and reheap down algorithms. They have analysed the algorithm and reached a logarithmic number of message complexity.

3 Background

A. Election Algorithm

Election algorithm is a special purpose algorithm, which is run for selecting the coordinator process among N number of processes. These coordinator or leader process plays an important role in the distributed system to maintain the consistency through synchronization. For example, in a system of client server, the mutual exclusion algorithm is preserved by the server process P_s, which is chosen from among the processes P_i where i = 1, 2, . . . ,N that are the group of processes which would use the critical region. Election Algorithm is needed in this situation to choose the server process among the existing processes. Eventually all the processes must agree upon the leader process. If the coordinator process fails due to various reasons, then immediately the election should happen to choose a new leader process to take up the job of the failed leader.

Any Election Algorithm should satisfy the following two properties [6].

(1) **Safety**: Any process P, has LEAD = NULL if it is participating in the election, or its LEAD = P, where P is the highest PID(process identifier) and it is alive at present.
(2) **Likeness**: All the processes should agree on the chosen leader P after the election. That is, LEAD = PID of the process P.

The Bully Election Algorithm [7] of Garcia Molina in 1982, elects the leader process uniquely which satisfies the safety and liveness requirements. Depending on a network topology, many algorithms have been presented for electing leader in distributed systems.

B. Ring Election Algorithm

Depending on a network topology, many algorithms have been presented for electing leader in distributed systems. The Ring Election Algorithm [8] is based on the ring topology with the processes ordered logically and each process knows its successor in an unidirectional way, either clockwise or anticlockwise.

When any process notices that the coordinator process has crashed, it creates an EL_MSG by inserting its own PID and sends the message to its successor. If the successor is also down, the message would skip that process and goes to the next process of the successor or to the next etc., till it reaches a process which is not dead, along the ring network. When the EL_MSG is received by any process, it adds its PID to the list in the message. Like this, all the available processes in the ring would insert their respective PID in the list. Finally, the EL_MSG comes back to the process which initiated the message and the process too would recognize that it only had initiated that message, by identifying its own PID in the list.

The Election initiator process analyses and finds the highest PID among the available processes converts the EL_MSG into CO_MSG and removes all the PIDs from the list except the highest PID. This CO_MSG message is circulated along the ring for one circulation to inform the running processes about who the new CO_ORDINATOR is. When this message is circulated once, it is discarded and the Election Algorithm ends here.

When the message comes back to the process that started it:

(i) The process sees its ID in the list.
(ii) It checks all the PIDs and decides the coordinator (the process with the hightest ID).
(iii) It changes the message type CO_MSG and enters the LEAD process in the message.
(iv) CO_MSG is circulated again.
(v) When it comes back to the process that started it, it gets discarded there.

Limitation of Ring Election Algorithm: Multiple election messages may happen in parallel when more than one process detects the failure of the coordinator process. This creates a lot of overhead of creating and servicing each and every election message. This causes heavy traffic and sometimes congestion in the network. In the best case, when a single process detects the crashed leader, N_i is obtained as follows:

$$N_i = n_e + n_c \tag{1}$$

where n_e refers to the number of EL_MSG and n_c refers to the number of CO_MSG. In the average and worst case, when all the N processes start the election message, N_i is obtained with $O(n^2)$ from the following equation

$$N_i = n(n_e + n_c). \tag{2}$$

This message complexity will drastically bring down the entire systems performance, as all the processes spend quite a lot of time in servicing these messages or processing these messages. In order to bring up the performance even during election, we need to have exactly one complete election happening instead of simultaneous redundant elections. All the other redundant election messages need to be killed. For solving this problem an improved election algorithm is proposed in the following section.

4 Message Efficient Ring Election Algorithm (MEREA)

In the previous section we have seen that in the average and the worst case scenarios, the number of messages that are exchanged between processes is high in the original Ring Algorithm. Therefore it imposes heavy traffic on the network. The proposed algorithm tries to intensively decrease the redundant election messages.

Assumptions

1. All the processes in the distributed group should have their clocks synchronized to each other. We have logical clock and physical clock synchronization algorithms namely Lamport's algorithm for the logical clock synchronization, Cristians and Berkeley algorithms for the real time clocks or the logical clocks.
2. The Network is perfect. (i.e. when any message is sent, it wont be lost/modified. It would reach the destination. If the destination process is alive, it can see the message which was sent to it. Here too, we have reliable primitives to keep the network perfect.

The proposed Message Efficient Ring Election Algorithm is given in Table 1. When process P realizes that the coordinator has crashed, it initiates the Message Efficient Election Algorithm by creating the EL_MSG. It inserts its ID, and the time

Table 1 Message Efficient Ring Election Algorithm (MEREA)

```
begin
    Step1: call Build EL_MSG
    Step2: for k: = 1 to n-1
                call update EL MSG
              send the EL_MSG to SUCC_i .
          endfor
    Step3: Destroy CO_MSG
end
```

of creation and circulates the message in the ring by throwing it to its immediate neighbor.

According to our assumption, all the processes in the group have their clocks synchronized, and so all the alive processes have the same time in their clock. When any process Q receives the EL_MSG, it reads its log, to check whether it has created any EL_MSG recently. Any one of the following 3 scenarios is possible.

i. Q would have initiated the EL_MSG before P.
ii. Q would have initiated the EL _MSG just after P.
iii. Q did not create any EL_MSG at all.

```
if (scenario (i) or (ii), then Q checks the Creation Time(T) of P_s EL_MSG)
    if (T(EL_MSG)_P > T(EL_MSG)_Q) then
        Q destroys the (EL_MSG)_P
  else
        Q adds its PID in (EL_MSG)_P and sends to its SUCC.
else
    Q adds its PID in (EL_MSG)_P and sends to its SUCC.
```

5 Mathematical Analysis

A. Best Case Analysis

Let N be the number of processes, In the Best case, only one process detects that the coordinator has crashed. Then, altogether, there will be 2n messages sent, one full circulation EL_MSG for a maximum of N-1 processes and one full circulation CO_MSG for a maximum of N-1 processes.

$$n_e + n_c = 2n \tag{3}$$

leading to O(n) message complexity. The Ring Election Algorithm and the our proposed Ring Election Algorithm have the same time complexity as in Eq. (1).

B. Worst Case Analysis

Ring Algorithms worst case message complexity is O(n^2) from Eq. (2). In our Algorithm, the worst case of the modified Election Algorithm is further divided into three more cases of Best, Average and Worst case.

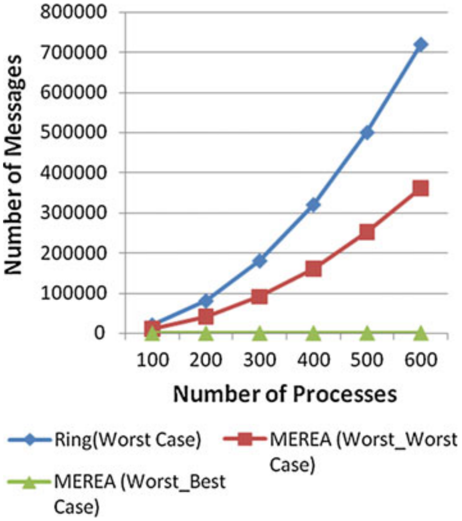
(1) *Best Case*: The best case is when, the processes initiate the Election in the anticlockwise direction according to time when the ring network is of clockwise direction. i.e. when the Election Algorithm gets invoked in the opposite direction of the ring according to time. The Election Messages get deleted in the very first go itself, that is when Process *i* send *s* election message to its immediate neighbor Process j and the message of *i* is killed by *j* because of time delay. Like this, except the very first or very earlier election message, all the other election messages would be killed by the immediate successor processes, which leads to the message complexity, O(n).

(2) *Average and Worst Case*: It is when the processes initiate the Election in the clockwise direction according to time when the ring network is of clockwise direction. i.e. when the Election Algorithm gets invoked along the direction of the ring according to time. Let Process *a* creates the first election message and passes it on along the ring. Then let its successor creates the Election message and passes to its successor and so on. In this case, only Process *a* can kill the election messages created by all the other processes. The duplicate election messages would take n1, n2, n3, …3,2,1 hops to reach Process *a*. The number of messages $\leq 1 + 2 + 3 + \cdots + n - 1$ which is of O(n^2).

6 Simulation Results

We simulated Ring election algorithm and MEREA in Java and the results are shown in Fig. 1. Every time we created number of processes such as 100, 200 and so on. The clock time of each process was synchronized. We kept 3 s message propogation time to reach from one process to another and 1 s to process the message. We made the process with the highest PID to be down which in turn invoked the Election activity automatically. We focused only on the worst case scenario because in the real time there is no possibility of having best case. Processes in the network with critical applications keep on sharing the common resources and exchange messages and this needs to be monitored and controlled by the coordinator. Worst case only is possible during these scenarios. In MEREA, the Worst case is again classified into two sub cases like Worst-Worst case and Worst-Best case. In Worst-Worst case, the message complexity is still O(n^2) but it drastically reduces the number of messages into one fourth of the messages as in Ring Algorithm. Our algorithm's Worst-Best case performance is O(n).

Fig. 1 Comparison of number of messages in Ring and MEREA



7 Conclusion

In this paper, we have proposed message efficient ring leader election algorithm, that reduces the number of messages in the election process. The duplicate EL_MSGs, which got destroyed before completion, the respective CO_MSG will also be removed. So, there will be only one CO_MSG. This is applicable for all the best, average and worst cases. N number of CO_MSGs are circulated in worst case of existing Ring Algorithm. No duplicate EL_MSG were destroyed in the existing algorithm, but our algorithm destroys them as early as possible by the help of clock time. These two factors give better performance in terms of time and reducing the messages. In future we would work on effective choosing of Cluster Head in Sensor Network.

References

1. Park SH (2011) A stable election protocol based on an unreliable failure detector in distributed systems. In: Proceedings of IEEE eighth international conference on information technology: new, generations, pp 976–984
2. Basu S (2011) An efficient approach of election Algori thm in distributed systems. Indian J Comput Sci Eng 2(1):16–21
3. Rahman MM, Nahar A (2009) Modified bully algorithm using election commission. MASAUM J Comput (MJC) 1(3):439–446. ISSN 2076 0833
4. Kim CY, Bauk SH (2006) The election protoco l for reconfigurable distributed systems. In: International conference on wireless networks (ICWN), pp 295–301
5. Kordafshari MS, Gholipour M, Jahanshahi M, Haghighat AT (2005) Modified bully election algorithm in distributed system WSEAS conferences. Cancun, Mexico, pp 11–14

6. Ben Ari C (2006) Principles of concurrent and distributed programming. 2nd edn. Pearson Education. ISBN: 10:032131283X, 13:978-0321312839
7. Garcia-Molina H (1982) Elections in distributed computing system. IEEE Trans Comput C-310:48–59
8. Andrew ST (2008) Distributed systems principles and paradigms. Tsinghua University Press, Beijing, pp 190–192

RAW: A Novel Reconfigurable Architecture Design Using Wireless for Future Generation Supercomputers

S. Viginesh, G. Vijayraghavan and S. Srinath

Abstract Current day Grand Challenge applications have conscripted scientists to move towards developing high performance computers. These applications are a part of varied scientific and engineering domains whose solution is impossible without the exercise of high performance computers. The need for faster computing in the order of exaFLOPs requires a type of architecture which complies with the power to performance ratio, satisfying the requirements of the end user. In this paper we present you a design for a node as well as cluster architecture of a high performance cum power aware supercomputer. The node architecture of the proposed design paradigm is made to be more application specific by the application of FPGAs or ASICs catering to the need of Grand Challenge Applications. A wireless routing technology is used for higher level communication i.e. inter application communication. Wireless technology makes architecture more reconfigurable catering to the communication complexity of different algorithms in the grand challenge applications. More over using wireless technology for higher level communication reduces the power consumption to a large extent thereby making this design paradigm suitable for varied class of grand challenge applications.

1 Introduction

GRAND challenge applications like brain modeling, climate modeling and protein folding possess high computational requirements. Millions of samples are required to work these applications. At this context, increasing the computational resources

S. Viginesh (✉) · G. Vijayraghavan · S. Srinath
Anna University, Chennai, Tamil Nadu, India
e-mail: vigineshsankararaman@gmail.com

G. Vijayraghavan
e-mail: vijayraghavan91@gmail.com

S. Srinath
e-mail: vecsrinath@gmail.com

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_82,

although can increase processing speed, but attention should also be put into developing architecture to promote parallelism in these applications as such a technique can massively boost performance with lesser computational resources. Apart from that, designing a supercomputer specifically for individual application will be cost prohibitive [1]. Node and cluster architecture of the supercomputer should suit simultaneous execution of multiple applications. In such architectures there is effective sharing of cost across multiple users and should necessarily be parallel execution and re-configurability in the next generation supercomputers, can be achieved by wireless communication in cluster architecture. To meet the demands, there should be a paradigm shift from the node architecture employed in current day supercomputers [2, 3]. The application of FPGAs as an accelerator to conventional general purpose processors can cater to proposed needs [4].

2 ASIC/FPGA Based Implementation of Node Architecture in High Performance Computers

Current day architectures [2, 3] used in high performance computers use a basic instruction set architecture based on x86-64. Solving a huge problem set in these processors require millions of instructions. This imposes a heavy workload to the operating system which enforces design engineers to evolve an architecture that balances the with a much lesser number of instructions. This results in the usage of hard wired accelerators supplementing conventional processors while evolving node architecture for grand challenge application execution in supercomputers [4].

To meet the demands for an efficient node architecture catering to the need of future generation high performance computers, we need to combine conventional processors with ASIC or FPGA units accelerating system performance. To supplement this idea let us look at the partially pipelined architecture of an NXN matrix multiplier unit (Fig. 1). The value of N, which is the number of parallel computations that can be done in single pipeline, can be chosen by suitable optimization techniques based on the customer's requirement, size of application input and performance parameters.

2.1 Odd Even Sorter

Parallel sorting of N elements can be done with the help of odd even sorting unit. The time complexity of odd even transposition sorting is O (bn2), where b = 1/2p2 and p represents the degree of parallelism. Hence it is an efficient sorting technique for high performance computing where parallelism is highly employed.

Fabricating ASICs for the purpose of using them as hardware accelerators, accelerating processor performance can be a seen as an expensive task. At this point one

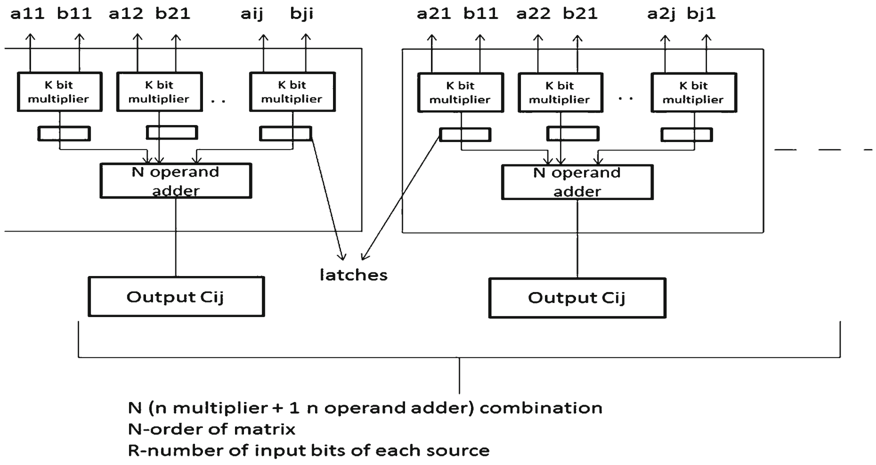


Fig. 1 Architecture of N X N matrix multiplication unit

can think of switching to reconfigurable architectures such as FPGAs as a substitute to ASICs. Due to the high programmability of FPGAs, they have shown high magnitude of improvement in performance power size and cost over conventional high performance computers. At this point one has to note that ASIC based architectures are much more performance efficient than that of FPGAs. We have chosen FPGAs to reduce cost and increase re-configurability.

Problems are of different types and hence no fixed architecture is likely to solve a large cross-section of the problem space very efficiently. Hence Reconfigurable architecture can best suit to high performance computing. At this point supplementing these individual clusters with FPGAs based on their functionality might be an apt measure to increase performance of the clusters. Apart from that, when we use FPGAs, the processing capabilities can be tailored to suit the needs of the algorithm. Moreover, the largest part of the die area of an FPGA is consumed by the configurable routing and this is important because in general, the biggest limiting factor of any implementation of a parallel algorithm is data passing.

3 Implementation of Wireless for Higher Level Communication

In the current trends of high performance computers, wireless clusters seem to overtake wired clusters but the only limitation that they possess is that the bandwidth restrictions of wired clusters is much lesser than that of wireless clusters. Our aim here in this work is to exploit the benefits of both wireless and wired system and construct a design for a cluster. Most of the network topologies used in high performance computers like 2D torus, 3D torus [2, 5] consumes power due to their intense communication links. Apart from that, all the nodes are not connected with each other which are highly inefficient as the number of computational nodes is scaled up. When wireless networks are employed, there is connection between each and every other node present in the cluster. A detailed analysis of parallel implementation of different algorithms in high performance computers can be done and the communication complexity can be modeled to determine the network traffic as well as communication pattern of different nodes in the cluster.

3.1 Organization of Nodes in RAW

Highly communicating nodes can be grouped together and placed geographically closer so as to decrease communication latency between thereby increase performance. Keeping this in mind, we propose a hierarchical system of communication where nodes organized in the first level of hierarchy will be employed with wired communication as their interaction is much more in magnitude rather than those connected in the second hierarchy where communication is by wireless networks. Hence we group computational nodes executing the same parallel algorithm into the first level of hierarchy which is known as the primary cluster. Algorithms which are a part of the application, given as input to supercomputers are initially parallelized and then given as input to the appropriate primary cluster. These parallel algorithms may highly communicate each other using wired communication topologies like 2D torus, 3D torus, mesh etc. The different nodes are connected by infiband, myrinet, gigabit Ethernet.

3.2 Primary Cluster Organization in RAW

There may also be dependencies across different algorithms in a high performance computing cluster. As the arrangement of each and every primary cluster in RAW is based upon the computational and communication complexity of individual algorithms, their network topology also is made to depict the same. Keeping this in mind, the communication between such nodes is very high. On the other hand the communication between different algorithms belonging to the same application given as input to the supercomputer is comparatively much lesser than that of within individual algorithm (Fig. 2).

Groups of primary clusters form secondary clusters. These secondary clusters communicate with each other by wireless networks. A possible solution for such communication can be by using MIMO (multiple input multiple output) based wireless communication. Multi user MIMO can be seen as the extended concept of space-division multiple access (SDMA) which allows a terminal to transmit (or receive) signal to (or from) multiple users in the same band simultaneously. This provides high bandwidth, increased capacity for channel and robustness. The network coverage of this technology is also sufficient to satisfy the needs of RAW based supercomput-

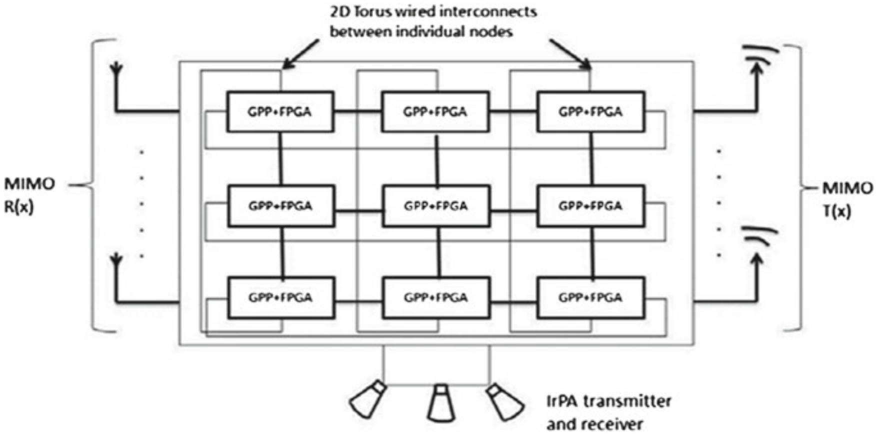


Fig. 2 Primary cluster architecture in RAW

ers. Even though MIMO based communication provides a good throughput, even higher throughput can be obtained by using IrDA (Infra-red Data Association) but this type of communication is limited by its range of propagation. With these principles in mind we evolve a hybrid wireless communication topology to exploiting the advantages of both MIMO and IrDA based wireless systems.

We group primary clusters belonging to the same application which communicate more frequently and place them geographically closer to each other which comes within IrDA's visibility. Most of the inter-algorithmic dependencies within the application can be solved by this method. Data transfers between primary clusters that are not in the range of IrDA are communicated by MIMO based wireless communication. By this we can achieve a high throughput of communication with less power consumption. An important point that has to be noted here is that even though the bandwidth offered by wireless communication is comparatively much lesser than that of wired communication, the fact that a single primary cluster can communicate to every other primary cluster present in the secondary cluster makes wireless implementation a much efficient technique and the amount of data that is being communicated thought the secondary cluster between primary cluster is much more in wireless than that of wired.

The Fig. 3 shows the individual primary cluster organization. Each and every primary cluster is supplemented with a set of transmitters and receivers as shown in the figure. These enable primary clusters which are not in IrDA's visibility range, communicate. Along with the set of transmitters and receivers, there are also a cells incorporated at each and every primary cluster [5]. These cellular transceivers transmit and receive infra-red radiation thereby enabling communication between them is very much possible. The entire organization of primary clusters together forming the secondary cluster is shown in the Fig. 3.

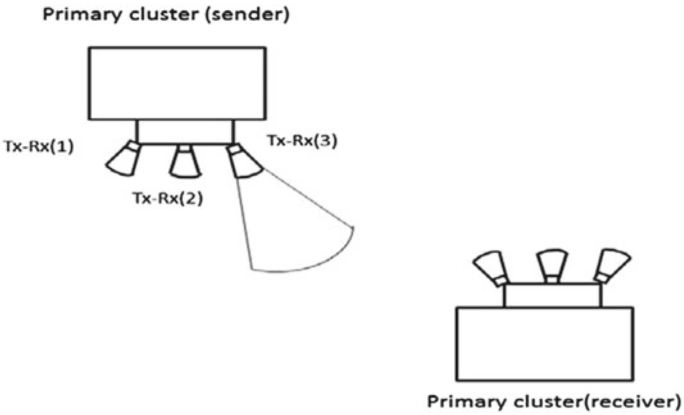


Fig. 3 IrDA based wireless communication between primary clusters

4 Mapping Applications to Cluster in RAW Based Supercomputers

With the advent of technologies to achieve exa-scale computing, researchers have started to adopt different strategies to parallelize algorithms to improve efficiency. Unlike the old trends, future computing systems are expected to run many diverse applications competing for shared resources. This diversity has also even led to the growth of many heterogeneous multi-cores at processor levels where parallelism has direct impact [4]. While prior research has tackled the problem of how to map threads belonging to a single application within a node, we need to evolve such a strategy that suits the need of the hour. There are many issues which could be solved using efficient application to node mapping and we list them down in this paper based on some heuristics (Fig. 4).

Firstly, algorithms those are part of an application which interacts within itself due to their high communication complexity should be mapped within the same primary cluster. This is because they communicate highly with each other and wired network topologies can best suit their requirements. Secondly, applications which are mapped closer to the shared or private memory source are seen to have high throughput in performance.

Before mapping application into their respective cores we should first convert them into a direct acyclic graph (DAG) form. Whenever mapping is done from the data repository to the cluster, appropriate FPGA/ASIC satisfying the finest grain thread of the algorithm for NXN matrix multiplication should be mapped (Fig. 5). Again data output of a particular FPGA will indeed act as the input for another. This gives the communication of data thought the path of the algorithm. Hence the communication complexity of the algorithm should match with the topology distribution of nodes in

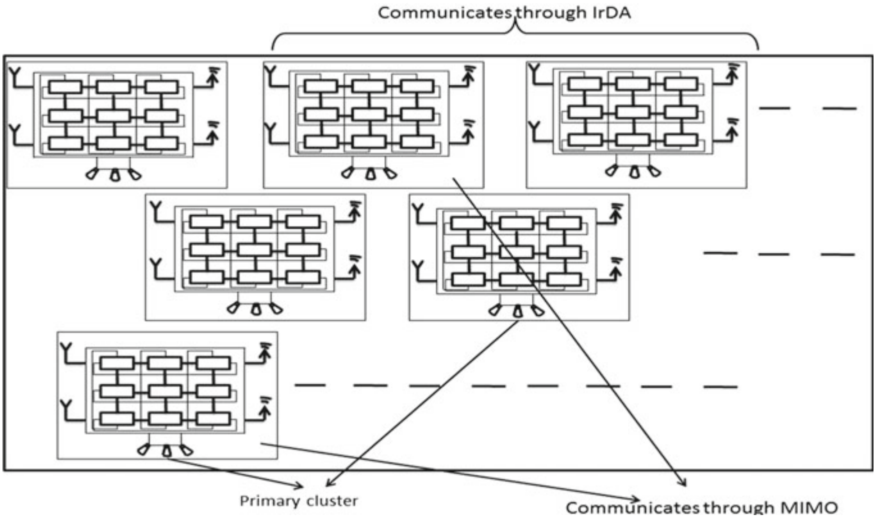


Fig. 4 Cluster design for RAW based supercomputers

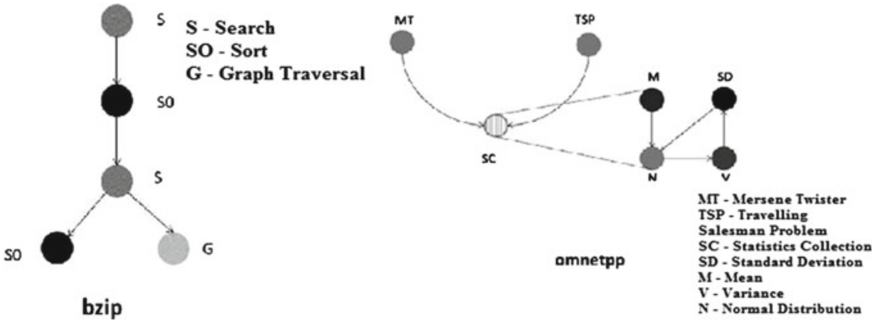


Fig. 5 DAG of bzip and omnetpp SPEC benchmark

the cluster. Similarly at the higher level of abstraction, different algorithms belonging to different applications should be mapped to the exact primary cluster. Again the communication across different algorithms should match to the network topology of different primary clusters. Wireless network topologies on the other hand suffice exactly to the above stated jurisdiction as they provide primary clusters with the highest level of re-configurability.

Our next task towards developing an application mapping strategy is to devise an algorithm to map different applications to a secondary cluster considering a set of heuristics. These strategies should correspond to simultaneous execution of multiple applications in the same supercomputer. Apart from that, an operating system perspective should also be provided to the design for scheduling various instructions that act upon different data. It is assumed that a network processor will be responsible for scheduling different inputs to the nodes executing the application. There is a scheduler that is associated with the network processor that maintains a queue of tasks that constitutes the DAG. Each and every task is scheduled to various primary clusters and their corresponding nodes based on their availability. Apart from these local queues also from an integral part of primary clusters which maintains tasks that are given as input to them by the network processor. They are again accommodated with mapping tables which contain information about different cores, tasks that the core has executed as well as those which are in its queue.

The different heuristics that are to be kept in mind while scheduling applications are as follows. Firstly, tasks should be mapped closer to dependent tasks as well as closer to the place where its related memory is present so that data and instruction latency can be avoided. Apart from that tasks should be mapped to nodes where the number of tasks in the queue waiting for that computational resource to get freed is lesser. By doing this inter as well as intra node traffic can be reduced to a great extent. Lastly, while mapping algorithms coming from different applications, the necessary condition that its dependent algorithms should be executed should always be true. Based the above heuristics tasks can be mapped be considering the following equation.

$$\boldsymbol{Map\,Criteria} = \boldsymbol{\varphi X}\left(\boldsymbol{\Sigma DAi}\right) + \boldsymbol{\psi}\left(\boldsymbol{QZ}\right)$$

DAi—distance between the considered node to be mapped and the ith dependent algorithm where i is the number of dependent algorithm.
QZ—is the size of the queue of the node which is considered to be mapped
Ψ and **φ**—are constants whose value lie between 0 and 1 whose value give weight to the heuristic that is to be considered.

5 Results and Analysis

The results given here are projected based on standard input workloads for benchmarking any architectural design. We have chosen SPEC (Standard Performance Evaluation Corporation) benchmarks and have converted into Direct Acyclic Graphs. These are the workloads that are given as an input to RAW.

Figure 6 shows the comparison between the numbers of instructions executed in a RAW based secondary cluster which includes FPGAs or ASICs as a part of its node architecture against node which contain only general purpose units. It has been seen that the numbers of instructions executed is much less in node architectures that includes FPGAs or ASICs. Apart from that, power consumption is reduced a lot as the number of memory fetch operation decreases.

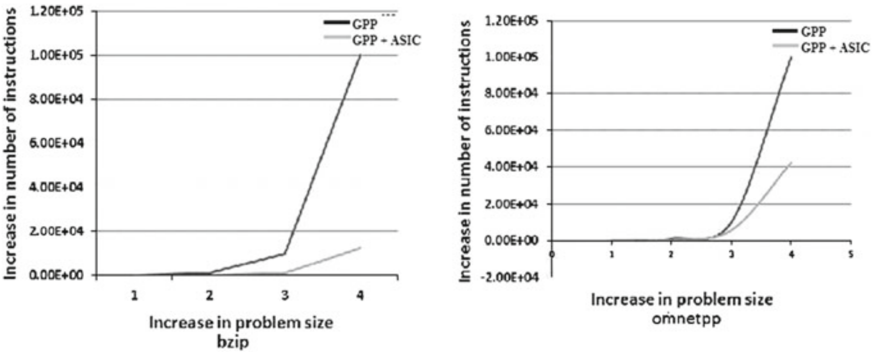


Fig. 6 Increase in problem size versus no. of instructions for bzip and omnetpp respectively

6 Conclusions

Era of Moore’s law can anytime come to an end. Considering technology scaling as a sole criterion for increasing node’s performance would lead a negative path to researchers for improving efficiency of supercomputers.

A practical solution to a problem of this kind would be using re-configurable node architecture. This may provide high performance due to scalability and re-configurability. Apart from that, power consumption can be done to a large extent by using a hybrid system containing wired and wireless networks. A supercomputer having such characteristics can be efficient for solving a varied class of grand challenge applications simultaneously catering to the needs of future generation supercomputers.

References

1. Watanabe T (1987) Architecture and performance of NEC supercomputer
2. Russel RM (1976) The CRAY-1 computer system. Cray Research Incorporation, Minneapolis
3. Langer KD, Vučić J (2010) Optical wireless indoor networks: recent implementation efforts
4. Chung ES, Milder PA, Hoe JC, Mai K (2010) Single-chip heterogeneous computing: does the future include custom logic, FPGAs, and GPGPUs? Computer Architecture Laboratory (CALCM). Carnegie Mellon University, Pittsburgh
5. Cvijetic N, Wilson SG, Brandt-Pearce M (2010) Optimizing system performance of free-space optical MIMO links with APD receivers

Interference Management Analysis of Double-ABBA and ABBA Quasi-Orthogonal Space Time Code

Mohammad Abu Hanif and Moon Ho Lee

Abstract In this paper we show the performance of Double-ABBA and ABBA scheme in with an interference management technique. We propose a system with partially cancel the interference which reduce the decoding complexity for higher order Multiple Input Multiple Output (MIMO) system. This interference cancellation scheme is called as interference management scheme. Therefore we use a group decoding technique proposed by Guo and Xia which is known as Partial Interference Cancellation (PIC) group decoding for Double-ABBA and ABBA Quasi-Orthogonal Space Time Coding. Finally we compare the performance of ABBA and Double-ABBA schemes.

Keywords Interference management · STBC · MIMO · QOSTC

1 Introduction

A space-time block coding (STBC)-oriented diversity scheme has been widely adopted in future wireless communication standards, such as 3GPP LTE, WiMax, etc. The STBC scheme, originally proposed by Alamouti in [1], achieves transmit diversity without channel knowledge. Even though Alamouti's STBC was originally designed for two transmit antennas and one receive antenna, this scheme has been generalized by Tarokh in [2] and extended to a system for four transmit antennas. There is also a lot of research to make STBC system work in the multi-user environment [2–4]. In order to achieve a high data rate, OSTBC is proposed in the literature

M. A. Hanif (✉) · M. H. Lee
Division of Electronic and Computer Engineering, Chonbuk National University,
Jeonju 561-756, Republic of Korea
e-mail: faisal_nkj@jbnu.ac.kr

M. H. Lee
e-mail: moonho@jbnu.ac.kr

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8_83,

by partitioning the transmit antennas into different groups and coding information symbols in each group with an independent OSTBC which is proposed in [3].

As a platform for high-speed wireless connectivity and networking, we need a reliable high data transmission rate. Thus, the transmit diversity requires more than one antenna at the transmitter side. We focus our presentation on Double-ABBA code proposed at [5]; the code is a simple permutation of 4 Alamouti codes of size 2 × 2 over 4 emitting antennas with spatial multiplexing rate 2. If utilizing conventional MMSE detection for DABBA, its need to do many inverse-calculations of 8 × 8 matrices, and this is infeasible in chip realization for its high complexity. Here, by utilizing the scheme, with Partial Interference Cancellation (PIC) group decoding proposed in [6], so it's reduced to low decoding complexity.

This work is organized as follows. In Sect. 2, we present the system model. In Sect. 3, simulation results are shown and the conclusion is given in the following section.

2 System Model

We consider a wireless system with Multiple Input and Multiple Output (MIMO) model. Where we assume, four users are transmitted data to a single receiver simultaneously. We consider the users and each receiver have multiple antennas, the channels are flat fading channels. We assume the wireless channel with Rayleigh distributed. Figure 1 shows the interference and desired signal for Double-ABBA scheme. In the

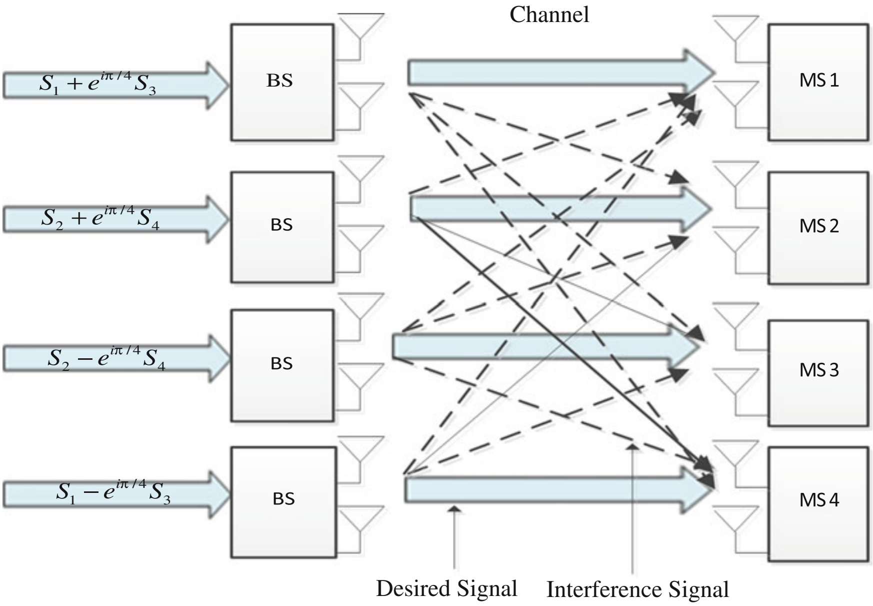


Fig. 1 Interference management analysis for multiple users

first part, we consider the Double-ABBA [7] code. As we know the Alamouti Space Time Code can be written as,

$$S = \begin{bmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{bmatrix}. \tag{1}$$

Therefore the DABBA code can be form as,

$$X = \frac{1}{\sqrt{2}} \begin{bmatrix} S_1 + e^{i\pi/4} S_3 & S_2 + e^{i\pi/4} S_4 \\ -e^{i\pi/4} S_4 + S_2 & -e^{i\pi/4} S_3 + S_1 \end{bmatrix}, \tag{2}$$

where, S_i, $i = 1, \ldots, 4$ is the Alamouti STC and $e^{i\pi/4}$ is define for the 45° rotation of. Now from (2) we can define X as follows

$$X = \frac{1}{\sqrt{2}}$$

$$\begin{bmatrix} x_1 + e^{i\pi/4} x_3 & x_2 + e^{i\pi/4} x_4 & x_5 + e^{i\pi/4} x_7 & x_6 + e^{i\pi/4} x_8 \\ -x_2^* - e^{-i\pi/4} x_4^* & x_1^* + e^{-i\pi/4} x_3^* & -x_6^* - e^{-i\pi/4} x_8^* & x_5^* + e^{-i\pi/4} x_7^* \\ x_5 - e^{i\pi/4} x_7 & x_6 - e^{i\pi/4} x_8 & x_1 - e^{i\pi/4} x_3 & x_2 - e^{i\pi/4} x_4 \\ e^{-i\pi/4} x_8^* - x_6^* & x_5^* - e^{-i\pi/4} x_7^* & -x_2^* + e^{-i\pi/4} x_4^* & x_1^* - e^{-i\pi/4} x_3^* \end{bmatrix} \tag{3}$$

So the received signal can be represent as,

$$Y = \frac{t}{\sqrt{2}} H X + w, \tag{4}$$

where, Y is the receive signals vector, H is the channel matrix, and w is the noise vector. Now taking the conjugate of y_2 and y_2 we can easily find the equivalent channel matrix, which is shown as,

$$\bar{H} = \begin{bmatrix} h_{11} & h_{12} & \alpha h_{11} & \alpha h_{12} & h_{13} & h_{14} & \alpha h_{13} & \alpha h_{14} \\ h_{12}^* & -h_{11}^* & \alpha h_{12}^* & -\alpha h_{11}^* & h_{14}^* & -h_{13}^* & \alpha h_{14}^* & -\alpha h_{13}^* \\ h_{13} & h_{14} & -\alpha h_{13} & -\alpha h_{14} & h_{11} & h_{12} & -\alpha h_{11} & -\alpha h_{12} \\ h_{14}^* & -h_{13}^* & -\alpha h_{14}^* & \alpha h_{13}^* & h_{12}^* & -h_{11}^* & -\alpha h_{12}^* & \alpha h_{11}^* \end{bmatrix} \tag{5}$$

where we put $a = e^{i\pi/4}$. The calculation of the pseudo-inverse of H would be very complex; this would be more complex for the matrix in higher dimension. Thus the decoding complexity increases very highly. Therefore we propose an interference management scheme for the DABBA STC which will describe hare. We first group the columns of equivalent channel matrix from (10) we have,

$$
\begin{aligned}
&g_0 = \left[h_{11}\ h^*_{12}\ h_{13}\ h^*_{14}\right]^T, && g_1 = \left[h_{12}\ -h^*_{11}\ h_{14}\ -h^*_{13}\right]^T,\\
&g_2 = \left[ah_{11}ah^*_{12}\ -ah_{13}\ -ah^*_{14}\right]^T, && g_3 = \left[ah_{12}\ -ah^*_{11}\ -ah_{14}\ ah^*_{13}\right]^T,\\
&g_4 = \left[h_{13}\ h^*_{14}\ h_{11}\ h^*_{12}\right]^T, && g_5 = \left[h_{14}\ -h^*_{13}\ h_{12}\ -h^*_{11}\right]^T,\\
&g_6 = \left[ah_{13}\ ah^*_{14}\ -ah_{11}\ -ah^*_{12}\right]^T, && \text{and } g_7 = \left[ah_{14}\ -ah^*_{13}\ -ah_{12}\ ah^*_{11}\right]^T.
\end{aligned}
$$

The grouping schemes are $I_0 = \{0, 5\}, I_1 = \{1, 4\}, I_3 = \{2, 7\},$ and $I_4 = \{3, 6\}$. Now the projection matrix is given by

$$
Q_{I_0} = \begin{bmatrix} h_{11} & h_{14} \\ h^*_{12} & -h^*_{13} \\ h_{13} & h_{12} \\ h^*_{14} & -h^*_{11} \end{bmatrix} \left(\begin{bmatrix} h^*_{11} & h_{12} & h^*_{13} & h_{14} \\ h^*_{14} & -h_{13} & h^*_{12} & -h_{11} \end{bmatrix} \begin{bmatrix} h_{11} & h_{14} \\ h^*_{12} & -h^*_{13} \\ h_{13} & h_{12} \\ h^*_{14} & -h^*_{11} \end{bmatrix} \right)^{-1} \tag{6}
$$

$$
\begin{bmatrix} h^*_{11} & h_{12} & h^*_{13} & h_{14} \\ h^*_{14} & -h_{13} & h^*_{12} & -h_{11} \end{bmatrix}
$$

So the projection matrix is,

$$
P_{I_0} = \frac{1}{|h_{11}|^2 + |h_{12}|^2 + |h_{13}|^2 + |h_{14}|^2}
\begin{bmatrix} |h_{11}|^2 + |h_{14}|^2 & h_{13}h_{14} - h_{11}h_{12} & -h_{11}h^*_{13} - h^*_{12}h_{14} & 0 \\ h^*_{13}h^*_{14} - h^*_{11}h^*_{12} & |h_{12}|^2 + |h_{13}|^2 & 0 & h_{11}h^*_{13} - h_{12}h^*_{14} \\ -h^*_{11}h_{13} - h_{12}h^*_{14} & 0 & |h_{12}|^2 + |h_{13}|^2 & h_{13}h_{14} - h_{11}h_{12} \\ 0 & -h_{12}h^*_{14} - h^*_{11}h_{13} & h^*_{11}h^*_{12} - h_{13}{*}h^*_{14} & |h_{11}|^2 + |h_{14}|^2 \end{bmatrix} \tag{7}
$$

The projection matrix for the second group as follows,

$$
Q_{I_1} = \begin{bmatrix} h_{12} & h_{13} \\ -h^*_{11} & h^*_{14} \\ h_{14} & h_{11} \\ -h^*_{13} & h^*_{12} \end{bmatrix} \left(\begin{bmatrix} h^*_{12} & -h_{11} & h^*_{14} & -h_{13} \\ h^*_{13} & h_{14} & h^*_{11} & h_{12} \end{bmatrix} \begin{bmatrix} h_{12} & h_{13} \\ -h^*_{11} & h^*_{14} \\ h_{14} & h_{11} \\ -h^*_{13} & h^*_{12} \end{bmatrix} \right)^{-1} \tag{8}
$$

$$
\begin{bmatrix} h^*_{12} & -h_{11} & h^*_{14} & -h_{13} \\ h^*_{13} & h_{14} & h^*_{11} & h_{12} \end{bmatrix}
$$

The projection matrix now become,

$$P_{I_1} = \frac{1}{|h_{11}|^2 + |h_{12}|^2 + |h_{13}|^2 + |h_{14}|^2} \begin{bmatrix} |h_{12}|^2 + |h_{13}|^2 & h_{13}h_{14} - h_{11}h_{12} & -h_{11}h_{13}^* - h_{12}^*h_{14}^* & 0 \\ h_{11}^*h_{12}^* - h_{13}^*h_{14}^* & |h_{11}|^2 + |h_{14}|^2 & 0 & h_{11}h_{13}^* - h_{12}h_{14}^* \\ -h_{11}h_{13}^* - h_{12}^*h_{14} & 0 & |h_{11}|^2 + |h_{14}|^2 & h_{13}h_{14} - h_{11}h_{12} \\ 0 & -h_{11}h_{13}^* - h_{12}h_{14}^* & h_{13}h_{14} - h_{11}h_{12} & |h_{12}|^2 + |h_{13}|^2 \end{bmatrix} \tag{9}$$

Similarly the projection matrix for the others grouping scheme for $I_3 = \{2, 7\}$,and $I_4 = \{3, 6\}$ are respectively.

$$P_{I_2} = \frac{1}{|h_{11}|^2 + |h_{12}|^2 + |h_{13}|^2 + |h_{14}|^2} \begin{bmatrix} |h_{11}|^2 + |h_{14}|^2 & h_{13}h_{14} - h_{11}h_{12} & -h_{11}h_{13}^* - h_{12}^*h_{14} & 0 \\ h_{13}^*h_{14}^* - h_{11}^*h_{12}^* & |h_{12}|^2 + |h_{13}|^2 & 0 & h_{11}h_{13}^* - h_{12}h_{14}^* \\ -h_{11}^*h_{13} - h_{12}h_{14}^* & 0 & |h_{12}|^2 + |h_{13}|^2 & h_{13}h_{14} - h_{11}h_{12} \\ 0 & -h_{12}h_{14}^* - h_{11}^*h_{13} & h_{11}^*h_{12}^* - h_{13}*h_{14}^* & |h_{11}|^2 + |h_{14}|^2 \end{bmatrix} \tag{10}$$

and,

$$P_{I_3} = \frac{1}{|h_{11}|^2 + |h_{12}|^2 + |h_{13}|^2 + |h_{14}|^2} \begin{bmatrix} |h_{12}|^2 + |h_{13}|^2 & h_{13}h_{14} - h_{11}h_{12} & -h_{11}h_{13}^* - h_{12}^*h_{14}^* & 0 \\ h_{11}^*h_{12}^* - h_{13}^*h_{14}^* & |h_{11}|^2 + |h_{14}|^2 & 0 & h_{11}h_{13}^* - h_{12}h_{14}^* \\ -h_{11}h_{13}^* - h_{12}^*h_{14} & 0 & |h_{11}|^2 + |h_{14}|^2 & h_{13}h_{14} - h_{11}h_{12} \\ 0 & -h_{11}h_{13}^* - h_{12}h_{14}^* & h_{13}h_{14} - h_{11}h_{12} & |h_{12}|^2 + |h_{13}|^2 \end{bmatrix} \tag{11}$$

Another scheme we consider is the co-ordinate interleaved criterion proposed in [8]. In briefly, from the orthogonal column of the equivalent channel matrix in [9] we can find the grouping scheme. For the first group $\tilde{I}_0$ [9] we have the projection matrix,

$$\tilde{P}_{I_1} = \frac{1}{\sqrt{8}|h_1|^2 + |h_2|^2 + |h_5|^2 + |h_6|^2} \begin{bmatrix} h_1 \left(h_1||_1^2 + |h_2|^2 + |h_5|^2 + |h_6|^2\right) \\ h_2^* \left(|h_1|_1^2 + |h_2|^2 + |h_5|^2 + |h_6|^2\right) \\ 0 \\ 0 \\ h_5 \left(|h_1|_1^2 + |h_2|^2 + |h_5|^2 + |h_6|^2\right) \\ h_6^* \left(|h_1|_1^2 + |h_2|^2 + |h_5|^2 + |h_6|^2\right) \\ 0 \\ 0 \end{bmatrix} \tag{12}$$

Similarly we can calculate the projection matrix for the orthogonal column group $\tilde{I}_1$[9] as,

$$\tilde{P}_{I_2} = \frac{1}{\sqrt{8}\left(|h_3|^2 + |h_4|^2 + |h_7|^2 + |h_8|^2\right)} \begin{bmatrix} 0 \\ 0 \\ h_3 \left(h_3||_1^2 + |h_4|^2 + |h_7|^2 + |h_8|^2\right) \\ h_4^* \left(|h_3|_1^2 + |h_4|^2 + |h_7|^2 + |h_8|^2\right) \\ 0 \\ 0 \\ h_7 \left(|h_3|_1^2 + |h_4|^2 + |h_7|^2 + |h_8|^2\right) \\ h_6^* \left(|h_3|_1^2 + |h_4|^2 + |h_7|^2 + |h_8|^2\right) \end{bmatrix} \tag{13}$$

Since the column we take here are orthogonal to each other, we can verify that $\boldsymbol{P}_{I_1}\boldsymbol{G}_{I_0} = 0$. For the grouping scheme $\tilde{I}_2 = \{1, 3\}$, $\tilde{I}_3 = \{2, 4\}$, $\tilde{I}_4 = \{5, 7\}$ and $\tilde{I}_6 = \{6, 8\}$. Now the optimal detection can be written as,

$$\hat{x}_{I_0} = \text{argmin}, ||P_{I_0} y_1 - \sqrt{SNR} P_{I_0} g_{I_0} \bar{x}_{I_0}||. \tag{14}$$

In the similar way of (14) we can find the other optimal symbols.

3 Simulation Analysis

In this section, we present the simulation results to show the performances of DABBA and ABBA scheme with the interference management scheme. For simulation we assume the channel is the Rayleigh fading channel. We also modulate and demodulate our symbol by using QAM modulation scheme. In Fig. 2 we can see the DABBA is achieving better BER performance than that of ABBA scheme. And partial interference cancellation decoding reduces the decoding complexity occurs in the higher order antenna diversity.

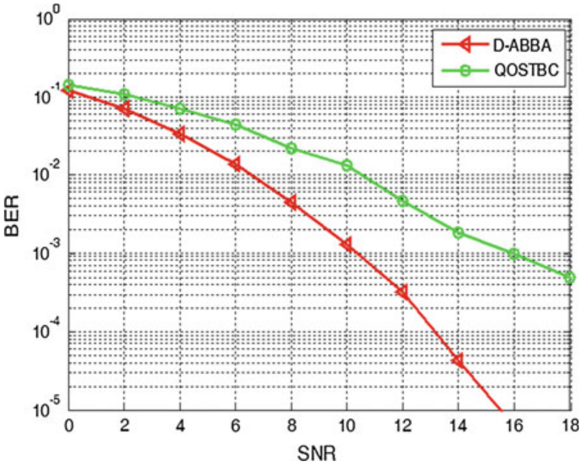


Fig. 2 SER analysis of QOSTBC with D-ABBA

4 Conclusion

In this letter, we present the interference management scheme for two well-known quasi-orthogonal space time code known as ABBA and Double-ABBA Quasi-orthogonal Space time-Block code. This scheme achieves full diversity because they are decoded with the partially cancelling the interference.

Acknowledgments This work was supported by World Class University (WCU) R32-2012-000-20014-0, NRF, Korea), BSRP (2010-0020942, NRF), Mid-career Researcher Program (National Leading Research Lab), MEST 2012-002521, and the second stage of Brain Korea 21 Project in 2012 (NRF), NRF Korea.

References

1. Alamouti SM (1998) A simple transmit diversity technique for wireless communications. IEEE J Sel Areas Commun 16:1451–1458
2. Tarokh V, Jafarkhani H, Calderbank AR (1999) Space-time block codes from orthogonal designs. IEEE Trans Inf Theory 45:1456–1467
3. Li G, Xia XG, Wu Y (2011) An optimal zero-forcing PIC group decoding for two-user layered Alamouti code. IEEE Trans Commun 59(12):3290–3293
4. Tarokh V, Naguib A, Seshadri N, Calderbank AR (1999) Combined array processing and space-time coding. IEEE Trans Inf Theory 45(4):1121–1128
5. Jorswieck E, Ottersten B, Sezgin A, Paulraj A (2008) Guaranteed performance region in fading orthogonal space-time coded broadcast channels. ERASIP J Wirel Commun Network 5:235–245
6. Guo X, Xia XG (2009) On full diversity space-time block codes with partial interference cancellation group decodin. IEEE Trans Inf Theory 55(10):4366–4385

7. Hanif MA, Lee MH (2012) Analysis of a group decoding scheme for double-ABBA Quasi-orthogonal STC. In: International conference on ICT convergence (ICTC)
8. Khan MZA, Rajan BS, Lee MH (2003) Rectangular co-ordinate interleaved orthogonal designs. In: IEEE global telecommunications conference, GLOBECOM '03
9. Hanif MA, Lee MH (2012) SER analysis of 8 × 8 QOSTC with less decoding complexity. In: International conference on cyber-enabled distributed computing and knowledge discovery (CyberC)

Author Index

N. Chaki et al. (eds.), *Computer Networks & Communications (NetCom)*,
Lecture Notes in Electrical Engineering 131, DOI: 10.1007/978-1-4614-6154-8,